TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Dolores Eamets 212099YVEM

Prevalence of high-risk human papillomavirus genotypes and their association with cervical cytology in Estonia: a population-based study

Master's thesis

Supervisor: Kaire Innos
PhD
Co-supervisor: Kadi Lubi
PhD

Tallinn 2023

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Dolores Eamets 212099YVEM

Suure onkogeense riskiga inimese papilloomiviiruse genotüüpide levimus ja nende seos emakakaela tsütoloogiaga Eestis: rahvastikupõhine uuring

Magistritöö

Juhendaja: Kaire Innos
PhD
Kaasjuhendaja: Kadi Lubi
PhD

Tallinn 2023

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Dolores Eamets

11.05.2023

Abstract

Background: Cervical cancer incidence and mortality in Estonia is the second highest in Europe. The leading cause of cervical cancer is the long-term human papillomavirus (HPV) infection. A population-based cervical cancer screening program was launched in Estonia in 2006. Women from the ages of 30 to 65 years old are invited to the screening on a five year-interval. The introduction of the HPV test in 2021 as the primary screening test enabled a population-based approach to studying the prevalence of high oncogenic risk HPV (hrHPV) genotypes. **Design:** Retrospective register-based cross-sectional study. **Objectives:** This study aimed to estimate the prevalence of hrHPV genotypes and their association with cellular changes in the cervix among women in Estonia for 2021. The secondary aim of this study was to analyse the screening data quality. **Methods:** A total of 37527 women aged 30-65 years, who had attended the organised screening in 2021, were included in the study sample. HrHPV positivity (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) was assessed using a vaginal swab (either physician or self-sampled). Data on screening test results were obtained from the Estonian Cancer Screening Register (ECSR). Completeness and accuracy of ECSR data were estimated through comparison with data obtained from laboratories. **Results:** The total hrHPV prevalence was 9%, highest in 30 to 35-year-old women. The most prevalent hrHPV genotype was HPV16 (2.5%). HPV16 individually and in combination with other hrHPV genotypes caused the most precancerous lesions in the cervix. The completeness of test result data in the ECSR was 70% for HPV and 83% for the LBC test. **Conclusions:** As previous studies have shown, the HPV prevalence was highest in younger women. The actual hrHPV prevalence in Estonia is most likely higher than 9%, because non-participants tend to be at a higher risk of developing cervical cancer. As expected, the hrHPV genotype with the most significant risk was HPV16. Human-centeredness needs to be implemented in the screening program by tailoring it to the needs of high-risk population groups. The switch to HPV test has improved data quality in the ECSR.

This thesis is written in English and is 65 pages long, including seven chapters, six figures, and five tables.

Annotatsioon

Kõrge onkogeense riskiga inimese papilloomiviiruse genotüüpide levimus ja nende seos emakakaela tsütoloogiaga Eestis: rahvastikupõhine uuring

Taust: Emakakaela vähki haigestumus ja suremus Eestis on Euroopas üks kõrgeimaid. Emakakaelavähi tekke suurimaks riskiteguriks on pikaajaline nakkus inimese papilloomiviirusesse (HPV). Eestis käivitati rahvastikupõhine emakakaelavähi sõeluuringuprogramm 2006. aastal. 30-65-aastaseid naisi kutsutakse sõeluuringule viieaastase intervalliga. HPV testi kasutuselevõtt esmastestina 2021. aastal võimaldas rahvastikupõhiselt uurida kõrge onkogeense riskiga HPV (hrHPV) genotüüpide levimust. **Disain:** Retrospektiivne registripõhine uuring. **Eesmärgid:** Uuringu eesmärgiks oli hinnata hrHPV genotüüpide levimust ja nende seost emakakaela rakuliste muutustega Eesti naiste seas 2021. aastal. Uuringu alaeesmärgiks oli analüüsida sõeluuringu andemete kvaliteeti. **Metoodika:** Uuringu valimisse kaasati 37527 naist vanuses 30-65 aastat, kes osalesid 2021. aastal sõeluuringul. HrHPV positiivsust (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) hinnati vaginaalsest sekreedist (proov võeti kas arsti poolt või ise). Andmed saadi vähi sõeluuringute registrist. Registriandmete täielikkuse ja täpsuse hindamiseks võrreldi neid laboritest saadud andmetega. **Tulemused:** HrHPV üldlevimus oli 9%, kõrgeim 30 ja 35 aastaste naiste seas. Suurima levimusega hrHPV genotüüp oli HPV16 (2,5%). HPV16 üksi ja kombinatsioonis teiste hrHPV genotüüpidega põhjustas kõige rohkem vähieelseid muudatusi emakakaelas. HPV testi tulemuste andmete täielikkus registris oli 70% ja LBC testi puhul 83%. **Järeldused:** Nagu varasemad uuringud on näidanud, oli HPV levimus kõrgeim noorimate naiste seas. Tegelik hrHPV levimus Eestis on suure tõenäosusega kõrgem kui 9%, sest mitteosalejatel on suurem risk haigestuda emakakaelavähki. Nagu eeldati, oli kõige suurema riskiga hrHPV genotüüp HPV16. Sõeluuringuprogramm peaks olema inimkeskne, kohandades seda vastavalt kõrge riskiga elanikkonnarühmade vajadustele. HPV testile üleminek on parandanud vähi sõeluuringute registri andmekvaliteeti.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 65 leheküljel, seitset peatükki, kuute joonist ja viite tabelit.

List of abbreviations and terms

AGC	Atypical glandular cells
AIN	Anal intraepithelial neoplasia
AIS	Adenocarcinoma in situ
ASC-H	Atypical squamous cells cannot exclude an HSIL
ASC-US	Atypical squamous cells of undetermined significance
ASIL	Anal squamous intraepithelial lesions
ECR	Estonian Cancer Registry
ECSR	Estonian Cancer Screening Registry
EHIF	Estonian Health Insurance Fund
HCD	Human-Centred Design
HIS	Health Information System
HPV	Human papillomavirus
hrHPV	High-risk human papillomavirus
HSIL	High-grade squamous intraepithelial lesions
HSP	Health service provider
ICD-10	International Classification of Diseases version 10
ID-code	National identification number
ITK	East Tallinn Central Hospital [In Estonian: *Ida-Tallinna Keskhaigla*]
LBC	Liquid-based cytology
LOINC	Logical Observation Identifiers Names and Codes
LSIL	Low-grade squamous intraepithelial lesion
LTKH	West Tallinn Central Hospital [In Estonian: *Lääne-Tallinna Keskhaigla]*
NIHD	National Institute for Health Development
NILM	Negative for intraepithelial lesion or malignancy
NOS	Not otherwise specified
Pap smear	Papanicolaou test
PERH	North Estonia Medical Centre [In Estonian*: Põhja-Eesti Regionaalhaigla]*

TEHIK	Health and Welfare Information Systems Centre [In Estonian: *Tervise ja Heaolu Infosüsteemide Keskus*]
TÜK	Tartu University Hospital [In Estonian: *Tartu Ülikooli Kliinikum]*
WHO	World Health Organization

Table of contents

List of figures

List of tables

1 Introduction

Despite the global efforts to reduce cervical cancer incidence with human papillomavirus (HPV) vaccination and population-based screening programs, cervical cancer is still the fourth leading cause of cancer morbidity and mortality in women [1]. Estonia has the second highest incidence and mortality rates of cervical cancer in Europe [2]. The leading cause of cervical cancer is long-term HPV infection [3].

Cervical cancer can be prevented with regular screening and HPV vaccinations [4]. Cervical cancer screening aims to detect precancerous conditions or tumours early in asymptomatic women to prevent cancer or begin early treatment [5]. Systematic population-based screening program can reduce the incidence rate of cervical cancer by 60-80% [6].

Estonia launched a population-based cervical cancer screening program in 2006 [7]. From 2006 until 2020, the primary screening test for the Estonian population-based cervical cancer screening was a Papanicolaou test (pap smear, cytological examination) which looks for abnormalities in the cervix. In 2021, the primary screening test was replaced by the HPV test, which looks for the presence of high-risk HPV (hrHPV) infection [8]. The introduction of the HPV test into the screening program enables the study of the prevalence of high oncogenic risk HPV genotypes in a population-based manner, using the birth cohorts in the 2021 cervical cancer screening target population [9], [10].

Population-based HPV prevalence estimates are a necessity for further planning of screening and additional research in Estonia [11]. As HPV testing has not been used for long as a primary test in screening, very few studies worldwide have examined the occurrence of cellular changes in the cervix in relation to different HPV genotypes [9], [10], [12], [13]. In Estonia, the connection between HPV genotypes and cellular lesions found on cytological examinations has never been studied before [14].

The overall goal of this study was to estimate the prevalence of hrHPV genotypes and their association with cervical cellular changes among women in Estonia for 2021.

In Estonia, the key stakeholder that measures the performance of the cervical cancer screening program is the Estonian Cancer Screening Register (ECSR). The register collects screening data electronically and periodically publishes key performance indicators [15]. Data for this study was obtained from the ECSR. As the quality of this study was influenced by the completeness and accuracy of the data in the ECSR, the secondary aim of this study was to evaluate the screening data quality.

This study consists of seven chapters. The introduction is followed by a background of the study topic. Then the study aims and objectives are introduced. Next, the methodology chapter gives an overview of the study design, time reference, population, and ethical considerations, as well as an introduction of the data collection and analysis methods used in this study. The results chapter is divided into three subchapters: the HPV and cervical lesions prevalence analysis results, followed by the data completeness and accuracy analysis results. The final chapters discuss and summarise the study's findings, conclusions, strengths, and limitations.

2 Background

This section provides an overview of the epidemiology of cervical cancer and the importance of cervical cancer screening. Additionally, this section introduces the population-based cervical cancer screening program in Estonia as well as the ECSR and screening test data standards. Lastly, this section discusses human-centeredness in screening.

2.1 The epidemiology of cervical cancer

Cervical cancer is a malignant tumour in the cervix [16]. According to the 10th version of the International Classification of Diseases (ICD-10), the codes C53 and D06 (cervix carcinoma in situ) are used in medical reporting to define cervical cancer [17]. When detected and treated early, cervical cancer is a curable disease, and with regular screening, it is also preventable [18].

Cervical cancer is the fourth most common cancer in women, with an estimated 604000 new cases diagnosed and 342000 deaths caused by cervical cancer recorded worldwide in 2020 [1]. The global burden of cervical cancer will continue to increase with the number of new cases diagnosed annually is projected to rise to 700000 and the number of deaths to 400000 in 2030 [18]. Most new cases of cervical cancer occur among unscreened women [19].

In Europe, over 66000 new cases of cervical cancer are diagnosed each year, and over 30000 women die from this disease annually [20]. The incidence of cervical cancer in Estonia is 27.4 per 100000 women, which is the second highest in Europe [2]. According to the Estonian Cancer Registry (ECR), during the last five years, an average of 147 women have been diagnosed with cervical cancer in Estonia every year, and an average 61 women die from this disease annually [21], [22]. Women in Estonia who do not attend regular screening, are lower educated, divorced or widowed, have inconsistent health insurance, and live in more remote regions are at a higher risk of developing cervical cancer [23]. To reduce cervical cancer incidence in Estonia, increased effectiveness of

the population-based screening program is need, especially for women who have higher risk status and are typically harder to reach [23].

2.1.1 HPV infection

The most significant risk factor for the development of cervical cancer is long-term HPV infection which is the most common sexually transmitted disease [24]. The link between cervical cancer and HPV was discovered in the early 1980s when HPV16 DNA was isolated from cervical cancer tissue [25]. It takes 10 to 15 years for cervical cancer to develop from the HPV-infected cervical epithelium [26]. The progression from HPV infection to cervical cancer can be prevented with early detection and treatment of precancerous lesions [27].

HPV infection spreads through sexual contact and does not generally cause any noticeable problems. HPV infection is widespread, at least half of women who have had sexual intercourse have been exposed to at least one strain of HPV [16]. For most women, the immune system fights off the HPV infection before it persists long enough to cause problems, but some strains of the virus are more harmful than others [24].

Over 200 types of HPV strains have been identified, of which approximately 15 genotypes (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 and 82) can cause cervical cancer as they can transform infected cells into malignant tumour cells [3]. These genotypes are categorised as hrHPV and can be detected with HPV testing [6]. A positive HPV test result indicates the presence of one or more hrHPV genotypes in the cells of the cervix [28].

Studies have shown that HPV-specific risks for developing precancerous cervical lesions differ by HPV genotype [9], [12], [13]. HPV16 and 18 are the most common hrHPV types associated with cervical cancer [12], [29]. A study on the Swedish population found that HPV genotypes 16, 18, 31 and 33 carry a 28% risk of developing high-grade lesions (CIN3+) [9]. A study on the Turkish population showed that HPV positives infected with HPV16, 35, 58 or 31 genotypes had the most cervical lesions found in the liquid-based cytology (LBC) test [13]. However, no similar studies have been conducted in Estonia yet, as the HPV test was made the primary screening test recently in 2021 [15].

The overall prevalence of HPV has been estimated to be 12% globally and 14% among women in Europe [30]. In the Dutch cervical cancer screening population, the prevalence of HPV was 8% in 2016 and 9.5% in 2021 [10], [31]. In Sweden, the prevalence of HPV among the 2021 cervical cancer screening population was 11% [32]. The most hrHPV positive results in the Dutch and Swedish cervical cancer screenings were found in the youngest age groups [31], [32]. In a Turkish cervical cancer screening population study, HPV was found in 3%, and the median age of HPV positive women was 42 ± 8.94 years [13].

The prevalence of the HPV virus has been studied in Estonia before, but in limited scope with a few studies using small study samples [11], [14], [33]. According to a study published in 2010, the prevalence of HPV among unvaccinated women in Estonia aged 18-35 was estimated to be 21%, and the hrHPV genotypes with the highest prevalence were HPV16, HPV53 and HPV66 [11]. In an HPV self-sampling study in Estonia, it was found that among 1903 women aged 37-62 years, the proportion of HPV positive results was 10% [33]. The prevalence of HPV in Estonia was also studied for 2020-2021 based on data collected in 2008, finding that the prevalence of HPV among women aged 30-33 was 22%, and the HPV strains with the highest prevalence were HPV16 and HPV56 [14].

2.2 Cervical Cancer Screening

To reduce the incidence of cervical cancer, 133 countries around the world have implemented national screening programs [34]. Systematic, population-level cervical cancer screening can reduce the incidence of cervical cancer by 60-80% [6]. Screening is an examination of healthy people without complaints or symptoms. The aim of cervical cancer screening is to identify abnormal cells which can evolve into cancer if left untreated or to diagnose cervical cancer at an early stage [5]. A woman screened regularly for cervical cancer has a lifetime risk of a 0.8% chance of developing this disease [16]. Women in Estonia who have not attended regular screenings have 2.35 times (confidence interval 1.85–2.98) higher risk of developing cervical cancer [23].

The World Health Organisation (WHO) has set seven screening and treatment approaches for a cervical cancer screening program that divides them into “screen-and-treat” and “screen, triage and treat” approaches [35]. In the “screen-and-treat approach”, the decision to treat is based on a positive test result on the primary screening test. In the

“screen, triage and treat approach”, the decision to treat is dependent on a positive test result on both the primary and the second screening tests with or without a histologically confirmed cancer diagnosis. According to the WHO Global Health Observatory data repository in 2021, 14% of the countries with a cervical cancer screening program were using the HPV test as their choice of screening method, cytology was used by 68%, and the rest of the countries were using visual inspection as the primary screening test [34]. In 2022, as part of Europe’s Beating Cancer Plan, a new approach to cancer screening was presented that recommends HPV testing for women aged 30 to 65 every five years to detect and prevent cervical cancer in the European Union [36].

2.2.1 Cervical cancer screening in Estonia

A pilot project for cervical cancer screening took place in Estonia in 2003, and a population-based screening program was launched in 2006 [7]. Women aged 30-65 are invited to cervical cancer screening at a five-year interval [15]. Excluded from the screening target population are women who have in the past five years had a diagnosis of vulvar, vaginal, uterine, or cervical cancer (ICD-10 codes C51–C55 and D06) [15].

In 2021, there were several organisational changes in the cervical cancer screening program compared to the previous years: uninsured women could participate in the screening on equal terms with insured women; the screening target group was expanded from 30-55 years of age to 30-65 years of age; the clinical follow-up in the case of a positive primary test changed and the HPV test was introduced as the primary test instead of a Pap smear [7], [15].

In 2021 and 2022, pilot and feasibility studies were conducted in Estonia that offered the HPV self-sampling option to women in the cervical cancer screening target group instead of the conventional screening test (HPV test) at a healthcare provider [33], [37], [38]. In 2022, women in Ida-Viru County could collect their self-sampling kits at the local pharmacy instead mail ordering [39]. During self-sampling, women collect a vaginal sample and mail it to the laboratory for HPV testing [40]. The first study results proved the feasibility and good acceptance of HPV self-sampling [33]. The second study, conducted in 2021, showed a 10% participation rate increase in the study target group with HPV self-sampling [37]. The study's initial results from 2022 show a 3% increase in participation in cervical cancer screening compared to the previous year and proved the feasibility and acceptance of self-sampling kit collection from the pharmacies [39].

2.2.2 Screening pathway in Estonia

The cervical cancer screening program is not just a single test but a pathway [41]. Figure 1 presents the Estonian cervical cancer screening program pathway.

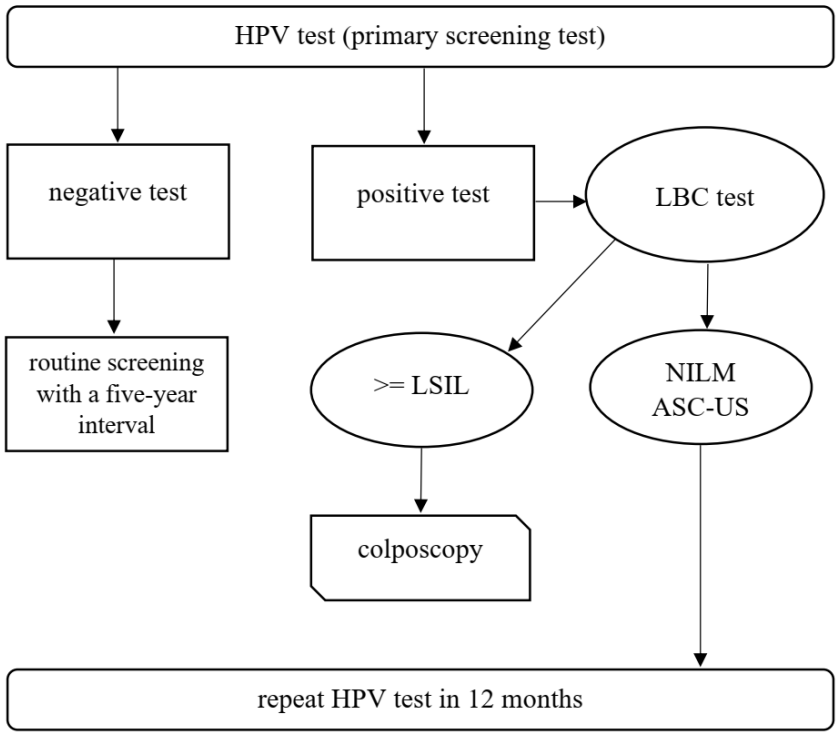

Figure 1. Screening logistics. Adapted from [15]. HPV=human papillomavirus; LBC=liquid-based cytology; LSIL=low-grade squamous intraepithelial lesion; NILM=negative for intraepithelial lesion or malignancy; ASC-US=atypical squamous cells of undetermined significance.

As shown in Figure 1, three different tests are used in the Estonian cervical cancer screening program: HPV test, LBC test and colposcopy. If the HPV test result is negative, the woman will be invited again to the cervical cancer screening after five years [15].

The primary test in the Estonian cervical cancer screening program is the HPV test [15]. The HPV test detects the presence of the hrHPV genotypes (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) in the sample material [15]. This test is taken during a pelvic exam where cells are collected from the surface of the woman's cervix with a brush [16]. Around 100 health care providers all over Estonia perform the HPV test as part of cervical cancer screening [42].

Once the cells are collected, the brush is rinsed in a vial of preservative solution and sent to the laboratory for analysis [15]. Currently, six laboratories analyse the primary screening tests in Estonia [43]. The laboratories are as follows: East Tallinn Central Hospital (ITK), Tartu University Hospital (TÜK), SYNLAB, West Tallinn Central

Hospital (LTKH), North Estonia Medical Centre (PERH) and, a recent addition as of autumn 2022, Pärnu Hospital.

When the HPV test result is positive, an additional cytological examination (LBC test) is performed, in which case a clinician-sampled test is taken from the same sample material [15]. In the case of a positive self-sampled HPV test, a woman must go to a health care provider separately to perform the LBC test because the test cannot be taken from the self-sampled material [33].

During a cytology test, the cells collected from the patient's cervix are tested in a laboratory to determine if they are precancerous [44]. The latest Bethesda system is used for reporting the LBC test results [15]. An essential component of the quality assurance of the Bethesda system is the sample material adequacy evaluation [45]. In the case of a satisfactory sample material, the test results are divided into two general categories: negative for intraepithelial lesion or malignancy (NILM) and epithelial cell abnormality [46]. According to the Bethesda system, epithelial cell abnormalities are divided into the following categories:

1) squamous cell abnormalities, including:
    a. atypical squamous cells of undetermined significance (ASC-US);
    b. atypical squamous cells - cannot exclude atypical squamous cells of undetermined significance (ASC-H);
    c. low-grade squamous intraepithelial lesion (LSIL);
    d. high-grade squamous intraepithelial lesion (HSIL);
    e. squamous cell carcinoma.
2) glandular cell abnormalities, including:
    a. atypical glandular cells (AGC) - endocervical, endometrial, glandular cells or not otherwise specified (NOS);
    b. atypical glandular cells favour neoplastic (AGC-FN) - endocervical or glandular cells;
    c. endocervical adenocarcinoma in situ;
    d. adenocarcinoma.

The cervical cytodiagnosis is reported based on the severity in a hierarchical manner as follows: carcinoma > HSIL > ASC-H > LSIL > AGC > ASC-US > NILM. If the LBC

test result is NILM or ASC-US, the women will be invited back to repeat the test in twelve months [15]. However, in case of a test result more severe than LSIL, the woman is invited to perform a colposcopy [15].

During a colposcopy exam, the cervix's surface is assessed, and a biopsy is performed if an abnormal tissue is found [16]. The resulting tissue sample is examined under a microscope to determine the presence of cancer [16].

2.3 Estonian Cancer Screening Registry

The ECSR was established at the National Institute for Health Development (NIHD) in 2015, with the aim to organise cancer screenings, analyse screening data, detect cancer early, evaluate the quality and efficiency of screenings, develop health policy and organise statistics and scientific research, including epidemiological research [47], [48]. The ECSR collects data on the tests connected with cervical, breast and colorectal cancer screenings and the data on treatment following the tests [48]. The ECSR's regular tasks include selecting the screening target group, creating referral letters for screening, sending out screening participation invitations and analysing and creating annual reports [49].

The ECSR was the first register in Estonia to collect data only in digital form [50]. Figure 2 presents the data exchange between ECSR and other databases.

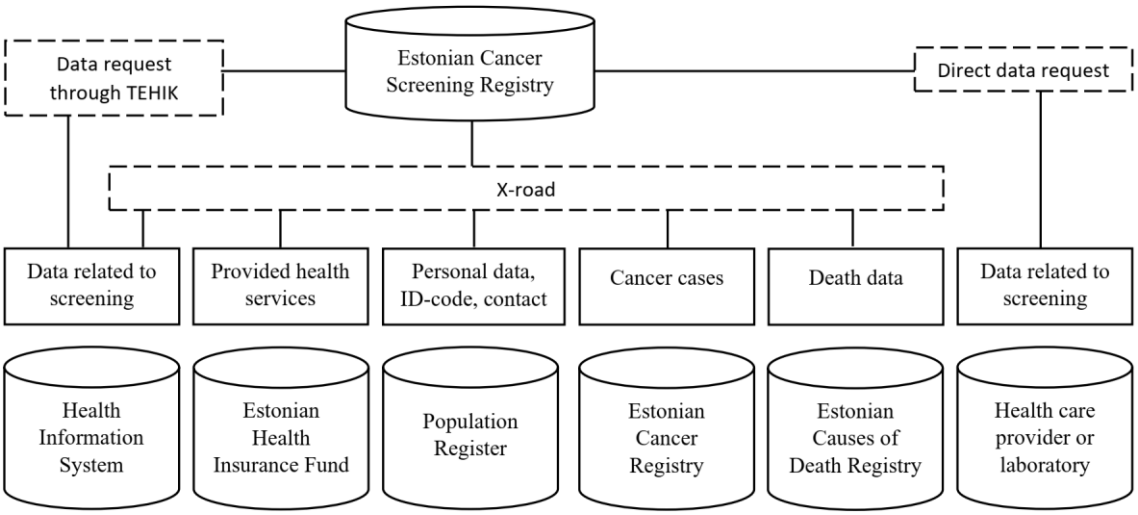

Figure 2. Databases and data exchange methods used by the Estonian Cancer Screening Register. Adapted from [51].

As can be seen in Figure 2, the ECSR functions as a digital register to which the data is obtained over the X-road data exchange layer from the Estonian National Health

Information System (HIS), the Population Register, Estonian Causes of Death Registry, ECR and EHIF. Direct data requests are also made to health service providers and laboratories [15]. The controller of the register is the National Institute for Health Development and, as of 2022, the processor of the register is the Health and Welfare Information Systems Centre (In Estonian: *Tervise ja Heaolu Infosüsteemide Keskus* - TEHIK) [49]. Since this change, the automatic data queries over X-road to HIS have not been updated, which impacts data quality [52]. These data queries are being replaced by a new data warehouse which should improve data quality in the register [52]. Until the data warehouse is ready for use, the ECSR has the right to make data requests to TEHIK, who can extract screening data directly from HIS [49].

The register publishes screening key performance indicators to the Health Statistics and Health Research Database three times a year: the cancer screening program's target population and invited to screening in January, detected cancer cases in cancer screening programs in May, and the cancer screening program's coverage by examination in August [53]. The register can publish only around 50% of the key performance indicators listed in the cervical cancer screening manual due to data capture and data quality problems in data collection from HIS [54].

Evelin Anion has studied data completeness in the ECSR in her master thesis, analysing Pap test result data in the register. Data was deemed complete if the register had received the Pap test results for women that participated in cervical cancer screening in 2016. Anion's study found that the screening data in the ECSR in 2016 was 51% complete [51]. However, according to the data disseminated at the Health Statistics and Health Research Database, the average cervical cancer primary test result data completeness from 2016 to 2019 was 38% [55].

2.3.1 HPV test data standards

The primary data source for ECSR is HIS, from which the register receives data on the primary cervical cancer screening test and, in case of detected pathology, also on additional examinations [48]. The health care providers and laboratories send screening data to HIS on the outpatient case summary or reply to the reference letter documents [56].

TEHIK has set standards on how to fill out these documents in the following manuals:

1. Instructions for filling out an outpatient discharge summary [In Estonian: *Ambulatoorse epikriisi täitmise juhend*] [57].

2. Instructions for filling out a referral response [In Estonian: *Saatekirja vastuse täitmise juhend*] [58].

Both documents state that the result field is compulsory for laboratory analysis if the sample material is adequate. HPV test analysis results should be filled as a numeric value, text value, or as according to the valid classification [57], [58]. The qualitative laboratory analysis result classifications are managed by the Estonian Society for Laboratory Medicine [59]. According to these classifications, the HPV test result data standards are as follows: negative (abbreviation N or numeric value 260385009), positive (abbreviation P or numeric value 10828004) and indeterminate result (abbreviation S or numeric value 280416009).

2.4 Human-centred design

The Human-Centred Design (HCD) originates from ergonomics, computer science and artificial intelligence. The International Organization for Standardization describes HCD as an approach to system design and development that applies human ergonomics and usability techniques to make interactive systems more usable [60].

In academic theory, Richard Buchanan, professor of design, management, and information systems at Case Western Reserve University, ties the practice of design to the promotion of human rights and human dignity to formulate that HCD is the affirmation of human dignity [61]. This means that in the practice of HCD, one should not only focus on creating a design for "users" or for usability but also research how the design can support and strengthen the dignity of human beings as they live their lives [61]. Joseph Giacomin, professor of HCD at Brunel University London, also sees that HCD is more than just the design's usability [62]. For Giacomin the best examples of design follow the HCD pyramid that at the base address questions about human physical, perceptual, cognitive, and emotional characteristics that are followed by progressively more complex, interactive, and sociological considerations [62].

The design consultancy firm IDEO uses HCD practices in their approach to design thinking. Tim Brown, the co-chair of IDEO, identifies design thinking as a human-centred

approach to innovation that integrates human needs, possibilities of technology, and business demands [63]. Brown places HCD and design thinking at the intersection of feasibility, viability and desirability of a design seen in Figure 3 [64].

HCD can be used to create a more person-focused health system. Researchers Kim Erwin and Jerry Krishnan believe that the focus should shift from helping people to fit the existing care delivery system to designing a system to fit the people where they live, work, learn, play, and receive healthcare [65]. In her doctoral dissertation, Julia Kramer proposed a framework that she adapted from Tim Brown model shown in Figure 3, which can be used to address global disparities in health access and challenges in global health equity [44].

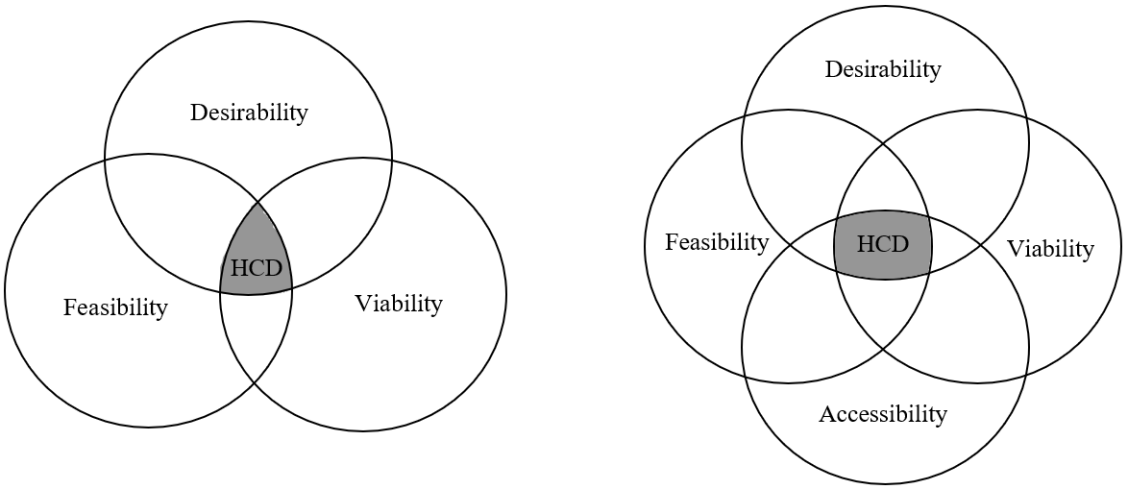

Figure 3. HCD frameworks by Tim Brown (left) and Julia Kramer (right). Adapted from [44], [64].

By applying Julia Kramer's HCD framework seen in Figure 3, a design can be called human-centred if it is functionally possible within the foreseeable future (feasible), sustainable (viable), makes sense to people and for people (desirable), and is accessible to the people who may interact with or benefit from the solution.

The Estonian Ministry of Social Affairs has set the patient's needs, involvement, and overall health outcome as a priority in the National Health Plan for 2020-2030. One of the key goals set in this plan is to lower inequalities in health care access. The National Health Plan states, "health care must be developed in a person-centred way and focused on preventing health problems" [66, p. 30].

2.4.1 Human-centeredness in screening programs

The prevalence of cervical cancer can be significantly lowered with HPV vaccination and regular screening [40]. As there is a limited time frame (12 to 15-year-old girls) during which a person can be vaccinated as part of the free national vaccination program in Estonia [67], there is still a significant population of unvaccinated women who are at risk of becoming infected with HPV and developing cervical cancer [44]. Therefore, the cervical cancer screening program must effectively complement the HPV vaccination campaign to lower the incidence and mortality of cervical cancer in Estonia.

The screening program will only make a substantial difference to population health if a sufficient number of women (70-80%) participate [68]. The participation in cervical cancer screening in Estonia is the lowest in Northern Europe [69]. The country's highest-ever participation rate was in 2021, with 51% of the eligible population covered by the primary screening test [70]. The screening participation rate is influenced by how convenient and acceptable the screening pathway is, beginning with informing the target group of screening eligibility up until follow-up and treatment where cancer or precancerous conditions are found [71]. Since the participation rate depends on the screening pathway, implementing HCD theory into the cervical cancer screening programme is necessary.

Cervical cancer screening should be accessible and convenient for participation regardless of the woman’s age, place of residence and socioeconomic status [8]. Women who do not have access to cervical cancer screening are less likely to identify cervical abnormalities before they become cancerous [44]. According to the evaluation of the screening pathway published by EHIF, women living in remote areas attend screenings less because going to the screening for them is more complicated. In contrast, women who live in the city do not attend the screening out of convenience [8]. For example, a woman living in the city might only attend the screening if the health care provider offering the screening happens to be on their daily trajectory. Another study in Estonia found differences in screening participation based on age, education level, marital status, and region of residence [72]. According to this study, older, single, and lower-educated women attended cervical cancer screening less [72]. Women from Viljandi had a low screening attendance compared to Hiiu County, where the screening attendance was the highest in Estonia [72].

HPV self-sampling can make cervical cancer screening human-centred. The feasibility and desirability of self-sampling among long-term screening non-attenders in Estonia were confirmed in a randomised feasibility study by Veerus et al. [33]. Implementing self-sampling in organised cervical cancer screening is viable, as it has been proven that HPV self-sampling is cost-effective and increases the participation rate [37], [73]. Lastly, HPV self-sampling is accessible as the test kit can be delivered to the woman's home, or they can receive it from their nearest pharmacy [38], [39].

Adding HPV self-sampling as an alternative screening method to clinician sampling may make the cervical cancer screening program design more human-centred. Over the past three years, HPV self-sampling studies have been conducted to make cervical cancer screening more accessible in Estonia [38]. The self-sampling option will continue to be offered to women invited to cervical cancer screening in 2023 on a project basis until it can be integrated into cervical cancer screening permanently.

Another way to implement HCD in cervical cancer screening is by making the screening pathway hrHPV-type specific. The risk for developing cervical cancer differs by hrHPV genotypes, such as HPV16 and 18 which cause more HSIL [29]. A hrHPV-type specific screening pathway would allow women who test positive for HPV types with higher risk to get to a diagnosis faster with no unnecessary wait time and extra tests. Such an option is especially relevant for women who test positive using a self-sampled HPV test as their primary screening test, as they would subsequently need to go to a clinic for an LBC test. For example, in Sweden, women with a positive test result for HPV16 or 18 and a normal cytological sample are offered new sampling after 18 months. In case of a type-specific persistence of HPV16 or 18 in the second HPV test, the woman is directly referred for colposcopy instead of follow-up cytology [74]. Such a method is not only human-centred, by creating a more efficient screening pathway based on the woman's needs. It could also lower the burden on health care providers as fewer unnecessary tests would be needed.

An HPV type-specific screening pathway has the potential to make the screening programme more accessible to high-risk groups and would therefore help to achieve the goals set in the National Health Plan. This study will examine hrHPV genotypes in association with cytology results to see if there is evidence for an HPV type-specific screening pathway in Estonia. The study results enable an evidence-based

recommendation for a screening pathway that can make Estonia's cervical cancer screening program more efficient in early discovery and preventing cervical cancer.

3 Aims and objectives

The problem statement of this research is:

HPV16 and 18 are the most common hrHPV types associated with cervical cancer [29], but although knowledge about the population-based prevalence of hrHPV types and their association with cervical lesions could support the development of a more efficient human-centred screening program, the hrHPV genotype association with cytology results has not been studied in Estonia.

The overall goal of this study is to estimate the prevalence of hrHPV genotypes and their association with cervical cellular changes among women in Estonia for 2021. As the quality of this study is influenced by the completeness and accuracy of the data in the ECSR, the secondary aim of this study is to analyse the screening data quality.

The specific objectives of this study are:

1) To assess the prevalence of HPV and hrHPV genotypes among women from the 2021 cervical cancer screening target population by age, region of residence, laboratory, method, and time of participation.
2) To examine the prevalence of cervical cellular changes among HPV positive women in relation to hrHPV genotype.
3) To evaluate the completeness and accuracy of the cancer screening register HPV and LBC test data, comparing data received from HIS and the laboratories.

The research questions for this study are:

1) What is the overall prevalence of HPV among women who participated in cervical cancer screening in 2021?
2) What is the overall prevalence of hrHPV genotypes among women who participated in cervical cancer screening in 2021?
3) What is the prevalence of HPV and hrHPV genotypes according to the 2021 cervical cancer screening participants' age and region of residence?

4) What is the difference in cervical abnormalities in connection to hrHPV genotypes?
5) What is the completeness and accuracy of the 2021 cervical cancer screening data in the ECSR?

Since the primary aim of this study is to describe the prevalence of HPV and hrHPV genotypes, hypotheses were not applicable for this study design [75].

4 Methodology

This section gives a detailed overview of the study population, time reference, data collection and analysis methods, as well as the ethical considerations of this master thesis.

4.1 Study design

This is a retrospective register-based cross-sectional study that uses a quantitative research design. The chosen research philosophy is positivism, as the research is based solely on facts, so that it can be objective. This study did not have direct contact with the study participants, as the data was collected from the ECSR without notifying the people in the study sample. The data received for this study from the ECSR is described in Appendix 2.

4.2 Time Reference

Data for this study was requested for all HPV tests taken from 01.01.2021 to 31.01.2022 by the women invited to the cervical cancer screening in Estonia. This time frame was chosen based on the ECSR method for calculating screening attendance. The chosen time frame includes all HPV tests taken at a health care provider and self-sampled tests. LBC tests included in this study were requested within six months following a positive HPV test result.

HPV test data quality in the ESCR is assessed in this study by comparing the cleaned dataset from 28.02.2023 to the initial register data from 10.01.2023. For the LBC test data completeness assessment, data were extracted from the ECSR on 15.03.2023.

4.3 Study population

The study population consisted of women who were invited to the cervical cancer screening in 2021 (year of birth 1956, 1961, 1966, 1971, 1976, 1981, 1986 and 1991) and had participated in the screening (had taken the HPV test at a health care provider or self-sampled test) in the period 01.01.2021-31.01.2022.

The screening target population includes women aged 30-65 whose residence indicated in the Population Register as of December 2020 was Estonia since that was when ECSR created the screening referral letter. Women who, according to ECR, HIS and EHIF data, had been diagnosed with vulvar, vaginal, uterine, or cervical cancer (ICD-10 codes C51–C55 and D06) from 2016 to 2021 were not invited to the screening. Therefore, by excluding these 462 women with a cancer diagnosis, 73803 women were invited to the cervical cancer screening in 2021 [76].

HPV prevalence was assessed among women invited to the cervical cancer screening in 2021 who had taken the HPV test at a health care provider or a self-sampled test in the study period (01.01.2021-31.01.2022). Therefore, the study sample consisted of 37527 women. Figure 4 presents the formation of the study sample.

The study population also includes women who had taken the HPV self-sampling test. In 2021 NHID, in cooperation with the EHIF, offered 26000 women an opportunity to choose between a regular screening visit and HPV self-sampling as part of a randomised pilot study for cervical cancer screening [37]. A total of 3541 women chose the HPV self-sampling kit instead of giving the test at a health care provider.

The HPV prevalence was analysed using HPV test data for all women in the study population. However, around 2% of the women in the study sample (n=773) had taken the HPV test more than once within the study time frame. For this study, only one HPV test per woman was used. For most women, the first HPV test taken within the study time frame was used for analysis. However, for 33 women, who had their first test result negative but a positive result for the following HPV test, the second test data was used for analysis.

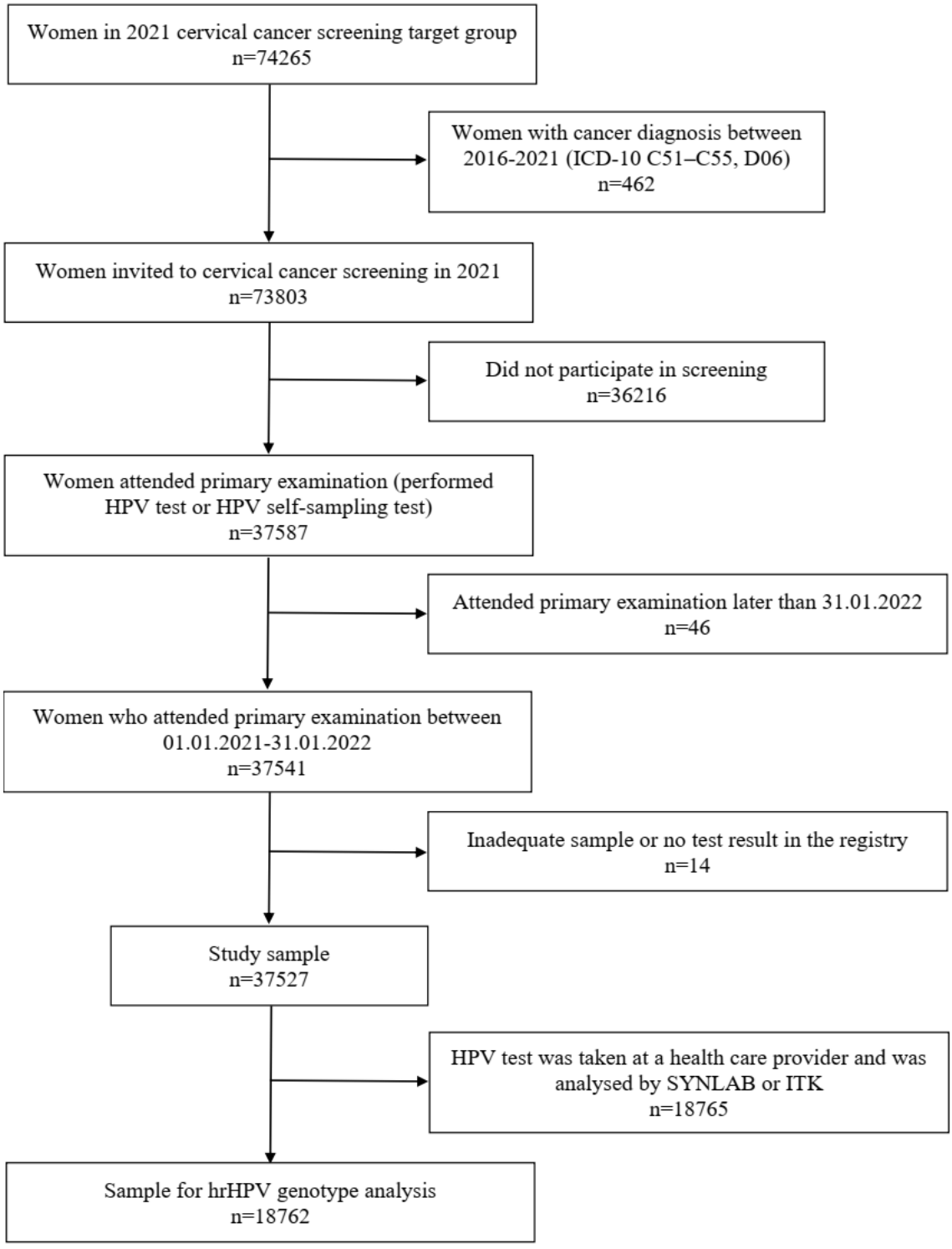

Figure 4. Study sample flowchart.

The cervical cancer screening manual sets requirements for the assays for hrHPV sequencing that can be used for the primary screening test. The laboratories are required to detect 14 hrHPV genotypes which are HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 [15]. Three different HPV sequencing assays were used by the laboratories that assessed the HPV tests in the 2021 cervical cancer screening. PERH, LTKH and TÜK

used the Alinity m hrHPV assay, which differentiates three individual hrHPV genotypes and two hrHPV genotype groups [77]. ITK used the Cobas® 4800 System, which differentiates two individual hrHPV genotypes and one hrHPV genotype group [78]. Lastly, SYNLAB used the PCR Multiplex and Luminex xMap technology, and in HPV test reporting, they differentiated one individual hrHPV genotype and two hrHPV genotype groups [79]. To achieve the study aims, only the HPV tests sequenced using Alinity m hrHPV assay were included for hrHPV genotype analysis because this assay differentiates most individual hrHPV genotypes.

Even though SYNLAB conducted the most HPV tests (38% of all HPV tests in the study sample), their HPV sequencing only enables the differentiation of the HPV16 genotype. For the HPV self-sampled tests, SYNLAB provided the exact genotype in case of a positive test (5% of all HPV tests by the study sample). The self-sampled tests analysed by SYNLAB were included in the study analysis as it was possible to format these HPV tests according to the Alinity m hrHPV assay method. The 12291 HPV tests, taken by a health care provider and analysed by SYNLAB, were excluded from further analysis. Also, 6474 HPV tests analysed by ITK were excluded from further analysis, as their genotyping method differed from the assay that TÜK, PERH, and LTKH used. Therefore, the study population for the hrHPV genotype analysis consisted of women whose HPV test was analysed by LTKH, PERH, TÜK, or the HPV self-sampled test analysed by SYNLAB (n=18762).

LBC test result was analysed only for women who had taken the LBC test following a positive HPV test. In total, 4530 women from the study sample took the LBC test, of which 2768 women took the LBC test following a positive HPV test. The women who had given the LBC test following a negative HPV test were excluded from further analysis. Out of 2768 women, 1289 HPV positive tests were analysed by LTKH, PERH, TÜK or the HPV self-sampled test analysed by SYNLAB. To ensure the validity of the analysis of cervical cellular changes in relation to HPV genotype, only the women with HPV tests using an assay that differentiated the most individual hrHPV genotypes were included. Therefore, the study population for the analysis of the prevalence of cervical cellular changes among HPV positive women in relation to HPV genotype consisted of 1289 women whose positive HPV test was analysed by LTKH, PERH, TÜK or the HPV self-sampled test analysed by SYNLAB.

4.4 Data collection

Data for this study was enquired from the ECSR. The complete dataset description can be found in Appendix 2. Since the author of this study is an employee at the ECSR, the data inquiries described in section 4.4.1 were done as part of the regular work tasks.

The study consists of data from the 2021 cervical cancer screening target population, that took the HPV test either at a health care provider or home (HPV self-sampling). The purpose of the subsequent data collection was to obtain the most comprehensive and accurate dataset of HPV test results for each woman. In the case of a positive HPV test, that also included a reference to the HPV genotype and the result of the additional study, the LBC test.

4.4.1 Data inquiries to other databases

As 2021 was the first year the new primary test was used, additional data queries were made to other national and health care institution databases for quality control purposes.

From the EHIF database, data was requested about HPV and LBC tests performed by the screening target population from 01.01.2021 to 31.06.2022. The data was requested based on the Estonian national identification number (ID-code) and the treatment procedure invoice code. Data were obtained from the EHIF database through a Power BI report.

Since the EHIF database does not provide information on the test result, additional data was requested from all laboratories that performed cervical cancer screening tests from 01.01.2021 to 31.06.2022. The laboratories are as follows: ITK, TÜK, SYNLAB, LTKH and PERH. Data was requested from those laboratories, consisting of the results of HPV and LBC tests performed for the cervical cancer screening 2021 target group based on person identifying ID-codes, the treatment procedure invoice codes, and Logical Observation Identifiers Names and Codes (LOINC). The laboratories sent data as an encrypted Excel spreadsheet. TÜK was not able to send LBC test data to ECSR.

An additional data request was made to HIS by TEHIK for the LBC tests that had been conducted for the screening target population from 01.01.2021 to 31.06.2022. The data was requested based on the screening target population ID-codes, the treatment procedure invoice codes, and LOINC. TEHIK sent data as an encrypted Excel spreadsheet.

Data received from EHIF, laboratories, and HIS was compared to the ECSR data using the Stata match function by ID-codes. The combined dataset was then further analysed in Excel. As 773 women had taken the HPV test multiple times and the ECSR receives data on one test from two different documents, there were multiple rows of data on one person. Therefore, the analysis time data field was used to connect data received from other databases and the existing data in the ECSR.

Data rows in the combined dataset were sorted into three different categories:

1) Data rows in ECSR which were complete and accurate (i.e., test result was filled correctly and was accurate in the register).
2) Data rows in ECSR which needed changing (i.e., the test result was not filled according to standard or was missing in the register).
3) Data rows missing from the ECSR (i.e., data was not available in the register).

Comma-separated value files were created for the data that needed to be uploaded to ECSR. Files were uploaded in multiple batches beginning with the data received from the laboratories, and the EHIF data was uploaded to the register only in a final step. After each file upload, the updated dataset was extracted from the register and then matched in Stata with the next received data file. This data-cleaning process lasted from January to February 2023 for HPV test data and from February to March 2023 for LBC test data. Data quality is assessed in this study by comparing ECSR data from before the data received from other databases was entered into the register (before the register data was cleaned) to the updated version.

4.5 Data analysis methods

The analyses were performed using Excel software and Stata 17 package.

4.5.1 HPV prevalence

The prevalence of HPV was studied in conjunction with the characteristics of the study sample (n=37527). The study sample characteristics were age, region of residence, HPV testing time, participation method, and laboratory where the test was analysed. The dependent variable was the HPV test result: positive or negative. Descriptive analyses of participants' characteristics and dependent variables were conducted as cross-tabulation

and presented in absolute numbers and percentages. HPV prevalence was calculated by dividing the number of women with positive HPV results by those in the study sample. A p-value of <0.05 was used to indicate whether findings were considered statistically significant.

For hrHPV genotype prevalence analysis, only HPV tests analysed by LTKH, PERH or TÜK and self-sampled tests from SYNLAB were included (n=18762). The HPV test results were categorised based on the Alinity m hrHPV assay, which individually identifies genotypes HPV16, 18 and 45, as well as reports on 11 other hrHPV types in two aggregates hrHPV (31, 33, 52, 58) and hrHPV(35, 39, 51, 56, 59, 66, 68) [77]. For the analysis, only one HPV test result per woman was used, which is why women who tested positive for multiple hrHPV genotypes were shown separately. The four combinations of multiple genotype positivity were HPV16 and 18; HPV 16, 18 and other types; HPV 16 and 45; HPV 18 and 45. Absolute numbers were presented for all HPV test results. The prevalence of hrHPV genotypes was calculated by dividing the number of women with each HPV test result by the number of women whose HPV tests were analysed by LTKH, PERH or TÜK and self-sampled tests from SYNLAB.

A descriptive analysis of the frequency and distribution of the various HPV types by age was performed. For distribution analysis of hrHPV genotypes among hrHPV positive women by age, only women who tested HPV positive and whose test was analysed by LTKH, PERH or TÜK and self-sampled tests from SYNLAB were included (n=1633). The HPV test results were presented in absolute numbers. The HPV test results and ages were conducted as cross-tabulation and presented in percentage, calculated by dividing the number of women with each result and age by the total number of women with that age. Findings with a p-value of <0.05 were considered statistically significant.

4.5.2 Prevalence of cervical cellular changes

The prevalence of cellular changes in the cervix was assessed among all women in the study sample who had taken the LBC test following a positive HPV test result (n=2768). The prevalence was calculated by dividing the number of women with each LBC test result by the total number of women included in the analysis.

A descriptive analysis of the frequency and the distribution of the LBC test results by hrHPV genotype was performed. The prevalence of cellular changes in the cervix in

association with the hrHPV genotype was assessed among women who had taken an LBC test, following a positive HPV test analysed by LTKH, PERH, TÜK or the HPV self-sampled test was analysed by SYNLAB (n=1289). The hrHPV genotypes were presented in absolute numbers. The LBC test results and hrHPV genotypes were conducted as cross-tabulation and presented in percentage, calculated by dividing the number of women with each result and the hrHPV genotype by the total number of women with that hrHPV genotype. A p-value was calculated for which <0.05 was considered statistically significant.

4.5.3 Data Quality

Screening data quality was assessed using the primary test data available in the register. To assess the HPV data quality in the ECSR database, two extractions were made. The two extractions only included women from the study sample. The first extraction was from 10.01.2023, and the second data extraction was made on 28.02.2023. Between the two extractions, the data in the register was cleaned and updated using data from the direct data inquiries to other databases, as previously explained in chapter "4.4.1 Data inquiries to other databases".

The two extractions were linked in Stata using the unique code that the register automatically generates for each data row. A person's ID-code was not used for linking because one woman could have multiple rows of data from different document types or because they had taken the primary screening test multiple times. The linked dataset was then exported and further analysed in Excel. The combined dataset consisted of 47022 rows of data. To know if the register had received HPV test results for all women who had participated in the 2021 cervical cancer screening, the duplicate data rows were removed. Using the filter function in Excel, the data rows were labelled based on whether the HPV test result was available or not in the 10.01.2023 and 28.02.2023 extractions. Then, duplicate rows of data were removed, leaving only one row of data per woman. If at least one data row for a woman contained a test result, the other data rows for this person were deleted. Finally, both extractions were compared. The data row was deemed of good quality if the HPV test result was accurate and complete, showing the proportion of accurate HPV test results in the ECSR.

Completeness was assessed using the qualitative laboratory analysis result classifications [59] and seeing if the analysis result data field was filled according to standard in the two

ECSR data extractions. Data was considered complete when the HPV result was filled according to standard. To calculate HPV result data completeness, the number of women who had their test result data available was divided by the study sample. This calculation was done separately for both extractions.

For the accuracy analysis, the two data extractions were compared to each other. The analysis result data was assessed in the extraction from 10.01.2023, and the extraction from 28.02.2023 was taken as a reference. Only data rows where the analysis result data field was filled were included in the accuracy analysis. Data was considered accurate if the result in the 10.01.2023 extraction had been filled according to the qualitative laboratory analysis result classifications set by the Estonian Society for Laboratory Medicine [59] and if it matched with the result in the 28.02.2023 extraction. To calculate result data accuracy, the number of women whose results were corrected (n=369) was subtracted from the number of women who had their result data available in the 10.01.2023 extraction (n=26252), and the difference was divided by the number of women who had their result data available in the 10.01.2023 extraction.

LBC test data completeness was assessed only after the data in the ECSR was cleaned and updated using data from the direct data inquiries to other databases. Data quality was not compared to an extract before the cleaning process began because ECSR collects data for follow-up tests with a two-year delay. In contrast, the primary test data is collected regularly in the ECSR. To calculate LBC result data completeness, the number of women who had their result data available was divided by the total number of women from the study sample whose LBC test data was in the ECSR. This analysis included only women who had taken the LBC test after a positive HPV test result.

Different HPV sequencing assays used by the laboratories were analysed. For all women in the study sample, the laboratory that analysed the HPV test included in this study was noted. Then, the *elhr.digilugu.ee* website was used to see which LOINC codes each of those laboratories used. For the ease of understanding the analysis results, the LOINC codes were then replaced with the name of the corresponding hrHPV genotype or genotype group. The LOINC codes and laboratories were conducted as cross-tabulation and presented in absolute numbers to see how many HPV tests in the study sample were done using the same HPV sequencing assays. Then the total number of tests done by each

laboratory was found, and the percentages of all tests included in this study analysis were calculated.

4.6 Ethical Considerations

The author of this study works as an analyst at the ECSR, which according to the legislation of the Estonian Cancer Screening Registry Statute, gave her the authority to use and process the necessary data from the ECSR database and enquire additional information regarding the performed screening tests from other national databases as well as from healthcare service providers and laboratories [49]. The prerequisite for conducting this study was the consent of the Research Ethics Committee of the University of Tartu (371/T-4, 21.11.2022).

In this study, no contact was made with the study sample as the data was enquired from the ECSR. Because the research subjects and their relatives were not contacted, there was no inconvenience or threat caused to the subjects. On the other hand, the inclusion of every subject who met the research criteria gave value to the study results since only a complete data set can provide an accurate overview of the population.

The study used personalised data, where there is a risk of breach of confidentiality and data leakage. Data collection, storage and analysis were carried out in accordance with the Personal Data Protection Act [80]. Data for this study was transferred, stored, and analysed solely on the NIHD servers to ensure data security.

4.6.1 Data Protection

The consent of the Research Ethics Committee of the University of Tartu was requested for the processing of personal data without the individual's consent in accordance with the conditions of § 6 (3) of the Personal Data Protection Act. The Personal Data Protection Act states that personal data can be used for scientific purposes without the consent of the data subject only if the data processing is unreasonably difficult to achieve without data-enabling identification, if there is a public interest in the study results, and if the data processing does not harm interests of the data subject [81].

This study used personalized data as data was acquired on two screening tests and descriptive data on the study population. It was only possible to link information about

one data subject based on their personal identifying ID-code. Personal data was necessary at the stage of data collection, for linking different databases, and assessing data quality. Data enquiries were made in accordance with set requirements. Strict security requirements were applied when transferring data, and personal data was encrypted for secure data exchange.

When processing personal data, the principle of purposefulness was considered [82]. Personal data was collected only to the extent necessary to achieve the defined study goals. To fulfil the objectives of the study, complete data was needed reflecting the results of the primary screening test and additional examination of the subjects.

This detailed dataset created in this study enabled the thorough analysis of the prevalence of HPV and hrHPV genotypes in the 2021 cervical cancer screening target group and the related precancerous changes in the cervix. Creating such evidence for Estonia was only possible using local data collected during screening, and the created data set is the basis for morbidity risk stratification. The study also helps to better assess the quality of cervical cancer screening in Estonia. Based on the results of this study, additional contributions can be made to a more effective cervical cancer screening organisation to ensure the early detection of precancerous changes and cervical cancer.

The data were processed on the secure NIHD server, located in Tallinn at Hiiu 42. The ECSR was queried based on the data subjects' ID-codes. After data on different screening tests and descriptive data on the study population were linked and the data quality was assured, the data was pseudonymized using the ECSR pseudonymisation key. This way, the data set could be linked back to register data where the pseudonymisation key was securely held and accessible only to the ECSR employees. For possible future needs of this study, the collected data set will be stored in a depersonalised form on the NIHD server indefinitely.

The study author ensures that all data protection principles were followed during this study. The collected data was used only for the purposes and scope of this research. Data processing in this study did not harm the data subject's interests because the output was a scientific generalisation. The study results were published as grouped indicators, which do not allow the identification of the subjects. In addition to this master thesis, the study's results will be published in scientific publications and in the ECSR reporting.

5 Results

This section provides results from HPV and hrHPV genotype prevalence, the prevalence of cellular changes of the cervix and data quality analyses.

5.1 HPV prevalence

Characteristics of the women who participated in the cervical cancer screening in 2021 and were included in the study sample are presented in Table 1.

Of the 37527 women included in the study sample, 3286 (9%) tested HPV positive. HPV prevalence was the highest in younger women (ages 30 and 35) and lowest in 55-year-old women. Differences were seen across counties, with Järva and Viljandi County having the highest HPV prevalence and Lääne County having the lowest HPV prevalence. Women who had taken a self-sampled HPV test had a higher HPV prevalence than those who gave the HPV test at a clinic. HPV prevalence also differed between laboratories.

Table 1. Characteristics of women participating in cervical cancer screening and their HPV test result, Estonia 2021.

			HPV positive result				
			No		Yes		
Variable	No.	col%	No.	row%	No.	row%	p-value[a]
Total	37527		34241	91.2	3286	8.8	
Age in 2021 (years)							<0.001
30	4609	12.3	3835	83.2	774	16.8	
35	5276	14.1	4650	88.1	626	11.9	
40	5061	13.5	4654	92.0	407	8.0	
45	5047	13.4	4655	92.2	392	7.8	
50	5170	13.8	4839	93.6	331	6.4	
55	4072	10.9	3849	94.5	223	5.5	
60	4387	11.7	4128	94.1	259	5.9	
65	3905	10.4	3631	93.0	274	7.0	
County							0.006
Harju	18962	50.5	17260	91.0	1702	9.0	
Hiiu	287	0.8	268	93.4	19	6.6	
Ida-Viru	2644	7.0	2446	92.5	198	7.5	
Järva	776	2.1	691	89.0	85	11.0	
Jõgeva	747	2.0	679	90.9	68	9.1	
Lääne-Viru	1572	4.2	1445	91.9	127	8.1	
Lääne	571	1.5	534	93.5	37	6.5	
Pärnu	2580	6.9	2369	91.8	211	8.2	
Põlva	613	1.6	563	91.8	50	8.2	
Rapla	892	2.4	822	92.2	70	7.8	
Saare	1043	2.8	949	91.0	94	9.0	
Tartu	4410	11.8	4031	91.4	379	8.6	
Valga	618	1.6	569	92.1	49	7.9	
Viljandi	973	2.6	866	89.0	107	11.0	
Võru	839	2.2	749	89.3	90	10.7	
Quarter of participation							0.024
2021 I	6297	16.8	5693	90.4	604	9.6	
2021 II	8590	22.9	7869	91.6	721	8.4	
2021 III	7458	19.9	6841	91.7	617	8.3	
2021 IV	13309	35.5	12147	91.3	1162	8.7	
2022 I	1873	5.0	1691	90.3	182	9.7	
Participation							<0.001
Clinic	34025	90.7	31223	91.8	2802	8.2	
Self-sampling	3502	9.3	3018	86.2	484	13.8	
Laboratory							0.002
ITK	6474	17.3	5852	90.4	622	9.6	
LTKH	3279	8.7	2979	90.9	300	9.1	
PERH	3237	8.6	2998	92.6	239	7.4	
SYNLAB	14075	37.5	12824	91.1	1251	8.9	
TÜK	10462	27.9	9588	91.6	874	8.4	

Figure 5 shows the hrHPV genotype prevalence in the study sample. The individual hrHPV genotype with the highest prevalence was HPV16 (2.5%). The prevalence of HPV18 was 1.1%, HPV45 was 0.4%, and other hrHPV genotypes was 6.2%. The highest prevalence was seen for the hrHPV genotype group that contained HPV35, 39, 51, 56, 59, 66 and 68.

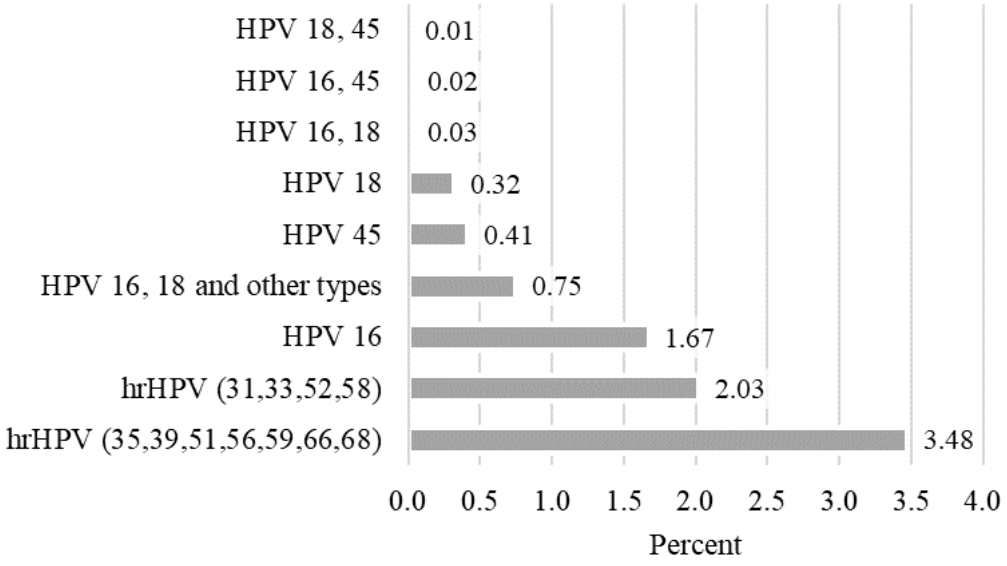

Figure 5. Prevalence (%) of hrHPV genotypes, Estonia 2021.

Table 2 shows the distribution of hrHPV genotypes by age. This analysis includes women whose positive HPV test was analysed by LTKH, PERH or TÜK or who took the HPV self-sampled test analysed by SYNLAB (n=1633).

Table 2. Distribution of hrHPV genotypes among hrHPV positive women by age, Estonia 2021.

		Age (years)							
		30	35	40	45	50	55	60	65
Variable	%	%	%	%	%	%	%	%	%
Total (No.)	1633	338	280	197	189	181	128	156	164
HPV genotype									
HPV 16	19.2	20.1	19.3	18.8	16.4	19.3	15.6	23.1	20.1
HPV 16, 18	0.3	0.9	-	-	0.5	-	-	-	0.6
HPV 16, 18 and other types	8.6	13.0	11.4	5.1	8.5	3.9	10.2	5.8	5.5
HPV 16, 45	0.2	0.6	-	0.5	0.5	-	-	-	-
HPV 18	3.7	3.8	2.9	5.1	3.2	1.1	2.3	5.1	6.1
HPV 18, 45	0.1	0.3	-	-	-	-	-	-	-
HPV 45	4.7	3.3	3.9	5.1	6.3	6.1	4.7	6.4	3.7
hrHPV (31,33,52,58)	23.3	25.7	23.6	29.4	23.3	18.2	21.9	17.3	22.6
hrHPV (35,39,51,56,59,66,68)	39.9	32.2	38.9	36.0	41.3	51.4	45.3	42.3	41.5

The individual hrHPV genotypes HPV16, 18, and 45 were prevalent for women of all ages. The individual hrHPV genotype with the highest prevalence for all ages was HPV16. HPV18 was the second most prevalent individual hrHPV genotype, and the least prevalent genotype was HPV45. The distribution of hrHPV genotypes differed significantly between age groups (p=0.034). For example, the hrHPV group (35,39,51,56,59,66,68) was the most prevalent in 50-year-old women, and the hrHPV group (31,33,52,58) was the least prevalent in 60-year-old-women.

5.2 Prevalence of cervical cellular changes

Around 16% of the women from the study sample (n=518) did not take an LBC test after a positive HPV test result. Figure 6 shows the prevalence of cytology (LBC) results for all women in the study sample who gave an LBC test following a positive HPV test result (n=2768).

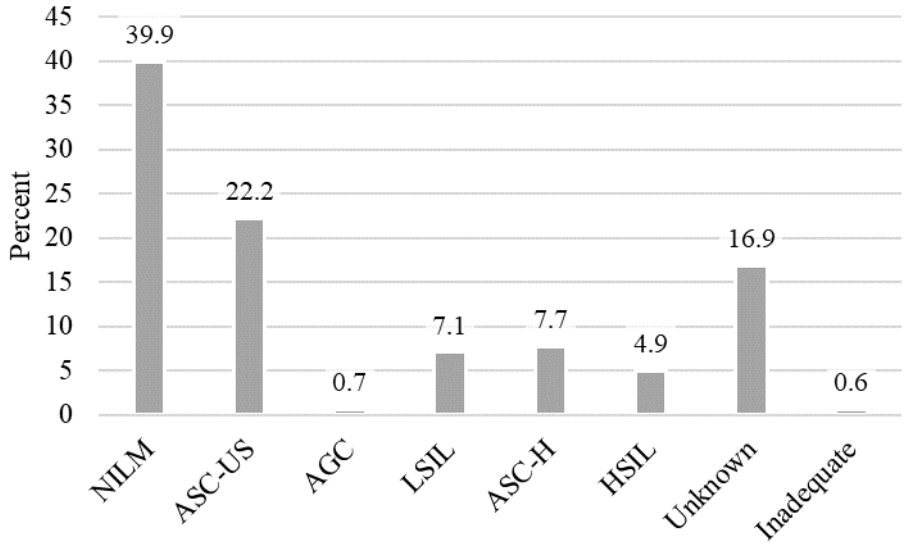

Figure 6. Prevalence of cervical cellular changes among HPV positive women, Estonia 2021.

No carcinomas were discovered from the cytology results following a positive HPV test in the study population based on the ECSR data. Approximately 40% (n=1104) of the women had a negative cytology result and were therefore invited to take a repeat HPV test after 12 months. Precancerous lesions (LSIL, ASC-H and HSIL) were found in nearly 20% of the women. For 17.5% of women, the test was inadequate, or no result was received by the ECSR.

Table 3 shows the relationship between the hrHPV genotype and the cytology (LBC test) result. This analysis includes women whose positive HPV test was either taken at a health care provider and analysed by LTKH, PERH or TÜK or who took the HPV self-sampled test and had given an LBC test (n=1289).

The distribution of cytology results differed significantly between hrHPV genotypes ($p<0.001$). HPV16 in combination with other hrHPV genotypes had caused the most precancerous lesions. HPV18 in combination with other hrHPV genotypes showed the second-highest proportion of LSIL, ASC-H and HSIL findings. Since the prevalence of HPV45 in combination with other hrHPV genotypes was minimal, there was not enough information to conclude that HPV45 combined with HPV16 or HPV18 causes the most precancerous lesions. HPV16 individually had caused the most HSIL, the hrHPV genotype group (31,33,52,58) the most ASC-H and hrHPV genotype group (35,39,51,56,59,66,68) the most LSIL. HPV18 and 45 individually had the lowest proportion of precancerous findings. However, HPV18 and 45 individually also had a low prevalence.

Table 3. Relationship between HPV positive result and cytology (LBC) result, Estonia 2021.

		Cytology result							
		NILM	ASC-US	AGC	LSIL	ASC-H	HSIL	Unknown	Inadequate
Variable	No.	%	%	%	%	%	%	%	%
Total	1289								
HPV genotype									
HPV 16	249	28.9	12.4	0.8	3.6	3.6	9.2	41.0	0.4
HPV 16, 18	5	-	-	-	-	40.0	-	60.0	-
HPV 16, 18 and other types	102	24.5	16.7	1.0	7.8	2.9	3.9	43.1	-
HPV 16, 45	4	25.0	-	-	-	50.0	-	25.0	-
HPV 18	55	36.4	21.8	-	3.6	1.8	-	36.4	-
HPV 18, 45	1	-	-	-	-	-	100	-	-
HPV 45	56	48.2	8.9	-	1.8	1.8	1.8	35.7	1.8
hrHPV (31,33,52,58)	302	34.1	16.2	0.7	3.6	6.6	5.0	33.8	-
hrHPV (35,39,51,56,59,66,68)	515	42.7	14.4	0.6	5.6	2.7	0.6	33.0	0.4

The p-value is <0.001.

5.3 Data Quality

Table 4 provides an overview of the data completeness and accuracy for the 2021 cervical cancer screening primary and secondary test data in the ECSR for the study sample. The analysis shows the data completeness and accuracy in ECSR for HPV test from 10.01.2023 and compares it to data from 28.02.2023 when the data from direct inquiries from the laboratories and HIS had been used to improve data quality in the ECSR. Secondary test data completeness was assessed using LBC test which was taken within six months after a positive HPV test. LBC test data was extracted from the ECSR on the 15.03.2023.

Table 4. Completeness and accuracy of HPV test results in the ECSR, Estonia 2021.

	HPV test		LBC test
Variable	10.01	28.02	15.03
Test data (No.)	37587	37587	2768
Result data available (No.)	26252	37578	2301
Corrected results (No.)		369	
Completeness* (%)	69.8	100	83.1
Accuracy** (%)	98.6		

* Completeness = number of tests results in the ECSR divided by the total number of test data.

** Accuracy = number of corrected test results subtracted from the number of test results in the ECSR on 10.01 divided by the number of test results.

The data completeness for HPV test data in the ECSR is around 13% lower than for LBC tests, however, the number of LBC tests is around 7% of the total number of HPV tests in the register. Even though the HPV test data accuracy in the register is close to 100%, the 369 women with corrected HPV test results in the register make up approximately 11% of all positive results in the study sample.

Table 5 shows the number of HPV tests analysed in each laboratory in the study sample.

Table 5. HPV test results reported by laboratories, Estonia 2021.

	ITK		LTKH		PERH		SYNLAB		TÜK	
HPV type	P	N	P	N	P	N	P	N	P	N
HPV 16 DNA	122	6352	53	3226	59	3178	-	-	210	10252
HPV 18 DNA	41	6433	19	3260	15	3222	-	-	50	10412
HPV 45 DNA	-	-	16	3263	11	3226	-	-	53	10388
hrHPV (31,33,52,58) DNA	-	-	94	3185	65	3172	-	-	242	10199
hrHPV (35,39,51,56,59,66,68) DNA	-	-	142	3137	107	3130	-	-	380	10061
hrHPV (31,33,35,39,45,51, 52,56,58,59,66,68) DNA	505	5969	-	-	-	-	-	-	21	-
hrHPV (16,18,31,33,35,39,45, 51,52,56,58,59,66,68) RNA	-	-	-	-	-	-	1251	12824	-	-
HPV 16 RNA	-	-	-	-	-	-	228	13847	-	-
HPV 18,45 RNA	-	-	-	-	-	-	105	13970	-	-
Total	6474		3279		3237		14075		10462	
%	17		9		9		38		28	

As can be seen from the table, SYNLAB and ITK collectively analysed over 55% of all HPV tests conducted in the study sample. The tests analysed by ITK do not differentiate HPV45 genotype. The tests analysed by SYNLAB only differentiate one individual hrHPV genotype, HPV16. The 1784 tests analysed by SYNLAB came from women using a self-sampled HPV test. From those 1784 self-sampled tests, SYNLAB provided the individual hrHPV genotype for all positive HPV tests.

6 Discussion

The aim of this study was to estimate the prevalence of hrHPV genotypes and their association with cervical cellular changes among women in Estonia to provide input for more efficient planning and tailoring of the screening program. The secondary aim was to evaluate screening data completeness and accuracy in the ECSR. This was a register-based cross-sectional study which used data from ECSR.

The HPV prevalence in Estonian women aged 30 to 65 was 9% in 2021. This study showed that HPV16 was the most prevalent individual hrHPV genotype in the 2021 Estonian cervical cancer screening target population. Women infected with HPV16 individually or in combination with other hrHPV types had the highest risk of developing precancerous lesions in the cervix. The completeness of HPV test result data in the ECSR was 70%, and accuracy was 99%. The completeness of LBC test result data in the ECSR was 83%.

6.1 Strengths and limitations

The main strength of the study was the large population-based sample size that makes this research unique in Estonia as it is the first study to assess HPV prevalence as well as hrHPV genotype prevalence in the screening population with such a large cohort.

There were also several limitations. Data quality analysis highlighted that the laboratories analysing the HPV tests for cervical cancer screening used three different genotyping methods, and not all data could be included in hrHPV genotype analysis. Only the laboratories which differentiated the most individual hrHPV genotypes were included. SYNLAB and ITK identified the least amount of individual hrHPV genotypes, which resulted in exclusion of around half of the study population from further analysis. Therefore, the difference between age groups compared to the hrHPV genotypes found in Table 4 may be random, as the number of women with each result was too small.

Another major limitation of this study was the completeness of cytology test data in the ECSR. The LBC test data completeness was mainly affected by the fact that TÜK could

not send LBC test data to ECSR. Therefore, the LBC test data without result information was received from EHIF, which resulted in 20% of unknown cytology test results in the study sample. Due to this, the prevalence of cytology results was not comparable to published data. The incompleteness of LBC test data also influenced the analysis of the relationship between hrHPV genotypes and cytology results. Because of incomplete data, the study results were unreliable in making an evidence-based recommendation to improve the efficiency of the cervical cancer screening program. For further studies, the LBC test data completeness in the ECSR must be improved.

6.2 HPV prevalence

This study found that the overall HPV prevalence in the study sample was around 9% which does not differ much from the results found in the HPV self-sampling study (10%) in 2020 [33]. Among the cervical cancer screening participants in 2021 in Sweden, the HPV prevalence was 11% [32], and 9.5% in the Netherlands [31]. In 2016 the HPV prevalence in the Dutch cervical cancer screening population was 8% [10]. The results from the Dutch screening programme are comparable to those found in this study among the participants in the Estonian national cervical cancer screening in 2021. Since the coverage by examination in the screening population was only 51% [70], the actual HPV prevalence in Estonia is most likely higher than 9% because women who do not attend regular screening tend to be at a higher risk of HPV infection [23].

HPV was more prevalent in younger women (30-35), consistent with a previous study conducted in Estonia which showed that the cervical cancer risk decreases with age [23]. However, the results found in this study were not contradictory to those of Kerli Reintamm's master's thesis, where the HPV prevalence in younger women in Estonia was studied. Her study found the HPV prevalence to be 22% among 30 to 33-year-old women compared to the 14% prevalence found in this study among 30 to 35-year-old women [14]. Another study in Estonia found the HPV prevalence to be 17% in 31 to 35-year-old women, which is closer to this study's findings [11]. The study sample sizes most likely influenced the difference in results since the sample sizes for the two studies were below 700 women [11], [14] compared to the 37527 women included in this study. Therefore, the prevalence in younger women found in this study is more accurate.

Studies have shown that women not attending regular screening have a higher risk of developing cervical cancer [19], [23]. This study found a significant difference in HPV prevalence in different regions of Estonia. Ida-Viru County has historically a low coverage by examination in cervical cancer screening, which should indicate a high HPV prevalence in this region [83]. The results showed a below-average (7.5%) HPV prevalence in Ida-Viru County. However, this was not an accurate representation of this region because less than 36% of the women living in Ida-Viru County who were invited to the screening in 2021 gave an HPV test [84]. In 2021, Viljandi County had the second lowest coverage by examination [83] and one of the highest HPV prevalence in this study. However, Saare County had the highest coverage by examination in 2021 [83] yet had an average HPV prevalence found in this study. Since for all Estonian counties the coverage by examination in women invited to screening was low (highest 60% and lowest 36%) [83], then it cannot be said with certainty, based on the results of this study, that there is a correlation between low screening attendance and HPV prevalence. To assess this correlation, HPV prevalence would need to be studied among women who do not regularly attend cervical cancer screening.

The study showed a higher HPV prevalence among women who gave self-sampled HPV tests (14%) than clinician-sampled HPV tests (8%). The first reason behind the difference in prevalence was that in 2021 SYNLAB detected two additional hrHPV genotypes, HPV33 and 82, for women who gave a self-sampled test [79]. Secondly, the PCR Multiplex and Luminex xMap technology used by SYNLAB had a higher sensitivity than the other assays used for HPV test analysis [85].

The difference in HPV prevalence was also found between laboratories that analysed the HPV tests. The difference can be explained by the fact that these laboratories use assays with different sensitivities and analyse tests for different regions in Estonia. ITK and LTKH mainly analyse HPV tests collected at Harju County, where the HPV prevalence was 9%, comparable to the prevalence found in those two laboratories. However, PERH, TÜK, and SYNLAB analyse HPV tests from across Estonia. To assess if there is a correlation between the laboratory and the region of residence, the laboratories would need to be contacted for further information on the logistics of the clinician-sampled tests.

The individual hrHPV genotype with the highest prevalence found in this study was HPV16 (2.5%) which, even though the prevalence differs, is consistent with the published

data. Studies conducted in Sweden and Turkey also showed that the hrHPV genotype with the highest prevalence in the screening population is HPV16 [12], [13], [32]. This study found that the prevalence of HPV18 was 1.1%, HPV45 was 0.4%, and other hrHPV genotypes was 6.2%. Prevalence of hrHPV genotypes has also been studied in Estonia before, but in younger women (ages 18 to 35), where the prevalence of HPV16 was the highest (6.4%), and the prevalence of both HPV18 and 45 was 0.6% [11]. This study also viewed the distribution of hrHPV genotypes by age and found that for all cervical cancer screening population ages, HPV16 was the most common individual hrHPV genotype, and the least common was HPV45.

6.3 Prevalence of cervical cellular changes

This study evaluated the HPV type-specific prevalence of cellular changes in the cervix among women who participated in the 2021 cervical cancer screening. Quantitative knowledge of hrHPV genotype-specific risks for precancerous lesions is necessary for estimating the effect of eliminating specific HPV types and the clinical benefits of screening for specific HPV types [9].

This study found that around 40% of the women in the study sample who had given a positive HPV test had a negative cytology result, which is less than is seen in screening results from Sweden and Norway (70%) [31], [32] as well as from studies conducted in Turkey (55-64%) [13], [73]. However, the prevalence percentages of cytology results cannot be reliably compared to published data because 17% of the LBC test results used for analysis in this study were unknown. In this study, cytological abnormalities (ASC-US, AGC, LSIL, ASC-H and HSIL) were found in 43% and precancerous lesions were found in 20% of the women.

Consistent with the published data, this study found that HPV16 alone, as well as in combination with other hrHPV genotypes, carry the most significant risk for the development of precancerous lesions [9], [12], [13]. A study in Sweden categorised HPV16 and 18 as the highest-risk oncogenic HPV types and HPV45 as a medium-risk oncogenic HPV type by evaluating the prevalence of precancerous lesions in women infected with HPV [9]. A study on the Turkish population found the most abnormal cytology results in terms of HPV subtypes in women who were infected with HPV16 individually or in combination with other hrHPV genotypes, being responsible for 75%

of all HSIL, 30% of all LSIL and 32% of all ASC-H cases [13]. Another study conducted in Turkey found that 20% of the women infected with HPV16 and 18 individually and in combination with a normal cervical cytology result had CIN2+ lesions found during a colposcopy-directed biopsy [12]. In contrast to published studies, this study did not find a significant risk of developing cervical cancer with HPV18 infection. However, nearly 40% of the women infected with the HPV18 genotype had their cytology result unknown. Therefore, the risk of HPV18 could not be assessed reliably in this study.

6.4 Data quality

The ECSR regularly monitors and evaluates the cervical cancer screening program [48]. ECSR was the first register in Estonia to gather data completely electronically [50]. The primary data source for the register is HIS, however, it is known that the quality of centrally collected data is incomplete [51], [52], [55]. This analysis showed that the register has reliable data on screening attendance (how many women attended screening). However, in many cases, the test result is not known, preventing quality assurance of the screening program. Collecting complete and accurate data in a timely manner is necessary for ECSR to measure the performance of the cervical cancer screening program to give their input on more efficient planning [52]. An efficient screening program is necessary to lower the incidence and mortality of cervical cancer in Estonia [6].

The completeness of the HPV test result data in 2021 was approximately 70% which is an improvement from when Pap-smear was the primary screening test. From 2015 to 2019, when Pap-smear was still the primary screening test, the data completeness in ECSR was below 50%, with the exception of 2016, where the cytology data completeness was concluded to be 51% [51], [55]. This shows that the implementation of HPV test in organised cervical cancer screening has improved data completeness in the ECSR. The study results also show that LBC data completeness (83%) in the register is better than for Pap-smear. However, it is important to note that the number of LBC tests done in 2021 was much lower than the number of Pap-smears done in a year as part of cervical cancer screening before 2021 [83].

Data inquiries were made by ECSR to the laboratories that analysed the HPV tests as part of the cervical cancer screening in 2021. The data received from the laboratories helped ECSR to get the HPV test data completeness to 100% and to improve data accuracy, as

for 369 women, the register had previously received a false negative HPV test result. Therefore, regular data exchange between laboratories and ECSR is necessary to attain quality data in the register. What is important to note is that data exchange between the laboratories and ECSR in the way that was done for this study is not a long-term solution, because it was a time-consuming process for both parties. A more viable solution would be an automated data exchange, as is the current practice used for data exchange between EHIF and ECR.

6.5 Conclusions

The study provided population-based estimates for overall and genotype-specific hrHPV prevalence in women of screening age group in Estonia as well as their association with cervical cellular changes. The study also estimated the completeness and accuracy of screening data.

The main conclusions of this study are:

1) The overall prevalence of HPV among women who participated in cervical cancer screening in 2021 was 9%. The actual hrHPV prevalence in Estonia is most likely higher than 9%, because non-participants tend to be at a higher risk of developing cervical cancer.

2) The overall prevalence of hrHPV genotypes among women who participated in cervical cancer screening in 2021 were 2.5% for HPV16, 1.1% for HPV18, 0.4% for HPV45 and 6.2% for other hrHPV genotypes.

3) The prevalence of HPV was the highest in younger women (30- and 35-year-olds) and lowest in 55- and 60-year-old women. The counties with the highest HPV prevalence were Järva and Viljandi and the lowest prevalence was in Hiiu County.

4) The most cervical abnormalities were found in women infected with HPV16 individually or in combination with other hrHPV types.

5) The completeness of HPV test result data in the ECSR was 70%, and accuracy was 99%. The completeness of LBC test result data in the ECSR was 83%. The swich to HPV test has improved data quality in the ECSR.

6) The study showed that the screening program should be tailored to the needs of high-risk population groups to make the screening pathway more human-centred. Human-centred screening programme design is essential since, according to the data in the ECSR, 58% (n=283) of the women, who had a positive self-sampled HPV test, did not go to a follow up LBC test. As more women choose self-sampling to participate in screening, it is necessary to provide them with a convenient way of continuing the screening pathway after a positive primary test for the screening to be most effective.

This is the first study in Estonia to describe the prevalence of HPV and hrHPV genotypes in a population-based manner in such a large cohort. Based on the results of this study, additional contributions can be made to a more effective cervical screening organisation to ensure the early detection of precancerous changes and cervical cancer.

6.6 Future directions

This study proved that the risk of developing precancerous cervical lesions is the highest among women infected with HPV16. However, this knowledge is insufficient to show the viability of an HPV-type-specific screening pathway. To make such a recommendation, further study is needed to assess if the HPV type-specific infection was still prevalent in the follow-up HPV test after 12 months following a normal cytology result.

Standardization of HPV genotype reporting across laboratories should be considered as it would lead to better comparability of primary screening test results and simplify the process of monitoring the entire screening pathway.

7 Summary

The aim of this thesis was to estimate the prevalence of hrHPV genotypes and their association with cervical cellular changes among women in Estonia and to analyse the screening data quality. Screening data for this study was obtained from the ECSR.

The screening data completeness in the ECSR was 70% for the HPV and 83% for the LBC test result data. HPV prevalence in women who participated in the cervical cancer screening in 2021 was 9%. Due to the low participation rate (51%), the HPV prevalence in Estonia is likely higher than this study found.

The individual hrHPV genotype with the highest prevalence was HPV16 (2.5%). This study found that women infected with HPV16 individually and in combination with other hrHPV genotypes had the most precancerous lesions in the cervix. Quantitative knowledge of the HPV type-specific risk of developing cervical cancer in the Estonian population is necessary to develop risk-stratified screening algorithms. These algorithms can help to create a more human-centred screening pathway to tailor to the needs of the high-risk population groups.

This study is unique in Estonia as it included a large study sample to detect the prevalence of HPV and its genotypes, as well as the relationship of hrHPV with abnormal cytology results.

Acknowledgement

I would like to express my deepest gratitude to my supervisors, Kaire Innos and Kadi Lubi, for their kindness and patience. But most importantly, their constructive feedback and support that kept me on course during the whole thesis writing process,

Special thanks to Elle Law for proofreading the thesis. I would also like to thank Birgit Saare, Eliise Leif, Steven Smit, and Siim Lassmann for the final revision and feedback on the thesis.

In addition, I would also like to thank Piret Viiklepp, Piret Veerus, Reeli Hallik and Madleen Orumaa for helping with the thesis idea development and revision of the ethics committee application.

I am also grateful to Kerli Linna, who provided me with the official English translations for the medical data fields.

I would also like to acknowledge the employees from ITK, PERH, TÜK, SYNLAB, LTKH, and TEHIK for providing the screening data needed for this thesis.

This work was supported by Estonian Research Council (grant no PRG722).

References

[1] H. Sung, J. Ferlay, R. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, "lobal Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians,* vol. 71, no. 3, pp. 209-249, 2021.

[2] ECIS - European Cancer Information System, "Estimates of cancer incidence and mortality in 2020, for all countries," 2023. [Online]. Available: https://ecis.jrc.ec.europa.eu. [Accessed 8 February 2023].

[3] I. Manini and E. Montomoli, "Epidemiology and prevention of Human Papillomavirus," *Annali di Igiene, Medicina Preventiva e di Comunità,* vol. 30, no. 4, pp. 28-32, 2018.

[4] M. Orumaa, K. Lutsar, J. Alloja, A. Uusküla, J. Jaal and M. Jürisson, "Sooneutraalse HPV-vastase vaktsineerimise kulutõhusus Eestis: tervisetehnoloogia hindamise raport, TTH58 [Cost-effectiveness of gender-neutral HPV vaccination in Estonia]," Tartu Ülikooli peremeditsiini ja rahvatervishoiu, Tartu, 2022.

[5] National Cancer Institute, "Cervical Cancer Screening," 13 October 2022. [Online]. Available: https://www.cancer.gov/types/cervical/pap-hpv-testing-fact-sheet. [Accessed 20 February 2023].

[6] K. T. Laisaar, T. Võrno, T. Raud, K. Jõers, D. Meigas-Tohver and R. A. Kiivet, "Inimese papilloomiviiruse (HPV) DNA-testi ja emakakaela tsütoloogilise uuringu (Pap-testi) võrdlus emakakaelavähi sõeluuringu esmastestina, TTH31 [Cost-effectiveness of the Estonian cervical cancer screening programme: Pap versus HPV testing]," Tartu Ülikooli peremeditsiini ja rahvatervishoiu instituut, Tartu, 2018.

[7] K. Innos, A. Baburin, R. Hallik and P. Veerus, "Rinna-, emakakaela- ja jämesoolevähi sõeluuringute tulemused Eestis [Breast, cervical and colon cancer screening results in Estonia]," 25 May 2022. [Online]. Available: https://ojs.utlib.ee/index.php/EA/article/view/18994. [Accessed 17 September 2022].

[8] H. Ental, H. M. Aavik and K. Lubi, "Emakakaelavähi sõeluuring–võimalus, mida ei tohiks jätta kasutamata. Teekonnaprojekti raport. [Evaluation of the cervical cancer screening pathway]," Tervisekassa, 2021.

[9] V. Smelov, K. Elfström, A. Johansson, C. Eklund, P. Naucler, L. Arnheim-Dahlström and J. Dillner, "Long-term HPV type-specific risks of high-grade cervical intraepithelial lesions: a 14-year follow-up of a randomized primary HPV screening trial," *International Journal of Cancer,* vol. 136, no. 5, pp. 1171-1180, 2014.

[10] C. Huijsmans, W. Geurts-Giele, C. Leeijen, H. Hazenberg, J. van Beek, C. de Wild, J. van der Linden and A. van den Brule, "HPV Prevalence in the Dutch cervical cancer screening population (DuSC study): HPV testing using automated HC2, cobas and Aptima workflows," *BMC Cancer,* vol. 16, no. 1, p. 922, 2016.

[11] A. Uusküla, M. Kals, L. Kosenkranius, L. A. McNutt and J. DeHovitz J, "Population-based type-specific prevalence of high-risk human papillomavirus infection in Estonia," *BMC infectious diseases,* vol. 10, no. 10.1186/1471-2334-10-63, p. 63, 2010.

[12] D. Altin and D. D. Keskin, "Relationship Between Human Papillomavirus (HPV), Cervical Smear Cytology and Colposcopy Directed Biopsy Results: 4 Year Experience of a University Hospital in Cervical Cancer Screening," *The Journal of Gynecology - Obstetrics and Neonatology,* vol. 19, no. 2, pp. 1295-1300, 2022.

[13] S. Findik, S. Findik, S. Abuoğlu, F. G. Cihan, H. Ilter and M. S. Iyisoy, "Human papillomavirus (HPV) subtypes and their relationships with cervical smear results in cervical cancer screening: a community-based study from the central Anatolia region of Turkey," *International journal of clinical and experimental pathology,* vol. 12, no. 4, p. 1391–1398, 2019.

[14] K. Reintamm, Kõrge riskiga HPV levimus ja sellega seotud tegurid sõeluuringut alustavate 30-33-aastaste naiste hulgas [High-risk HPV prevalence and the connected factors among 30-33-year-old women starting screening], Tartu: Tartu Ülikool, 2022.

[15] Estonian Health Insurance Fund, "Emakakaelavähi sõeluuringu tegevusjuhend [Cervical cancer screening manual]," 2020.

[16] L. C. Hartmann and C. L. Loprinzi, Mayo Clinic Guide to Women's Cancers, Rochester: Mayo Clinic, 2005.

[17] World Health Organization, "International statistical classification of diseases and related health problems (11th ed.)," 2019. [Online]. Available: https://icd.who.int/. [Accessed 18 February 2023].

[18] World Health Organization, "Global strategy to accelerate the elimination of cervical cancer as a public health problem," World Health Organization, Geneva, 2020.

[19] A. Subramaniam, J. M. Fauci, K. E. Schneider, J. M. Whitworth, B. K. Erickson, K. Kim and W. K. Huh, "Invasive Cervical Cancer and Screening: What are the Rates of Unscreened and Underscreened Women in the Modern Era?," *Journal of lower genital tract disease,* vol. 15, no. 2, pp. 110-113, 2011.

[20] World Health Organization, "Seventy-second Regional Committee for Europe: Tel Aviv, 12–14 September 2022: roadmap to accelerate the elimination of cervical cancer as a public health problem in the WHO European Region 2022–2030," in *Regional Committee for Europe, 72nd session*, Tel Aviv, 2022.

[21] Health Statistics and Health Research Database, "PK10: New cases of malignant neoplasms by specified site, sex and age group," 1 September 2022. [Online]. Available: https://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__02Haigestumus__04PahaloomulisedKasvajad/PK10.px/. [Accessed 8 February 2023].

[22] Health Statistics and Health Research Database, "SD21: Deaths by cause, sex and age group," 8 June 2022. [Online]. Available: https://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__01Rahvastik__04Surmad/SD21.px/. [Accessed 8 Feburary 2023].

[23] O. Nõmm, P. Veerus, M. Orumaa and K. Innos, "Effect of Pap-smear and sociodemographic factors on cervical cancer risk in Estonia: A population-based case-control study," *Cancer Epidemiology,* vol. 80, 2022.

[24] American Cancer Society, "Cervical Cancer Causes, Risk Factors, and Prevention," 3 January 2020. [Online]. Available: https://www.cancer.org/cancer/cervical-cancer/causes-risks-prevention/risk-factors.html. [Accessed 8 February 2023].

[25] M. Dürst, L. Gissmann, H. Ikenberg and H. zur Hausen, "A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions," *Proc Natl Acad Sci U S A,* vol. 80, no. 12, p. 3812–3815, 1983.

[26] World Health Organization, "Cervical cancer screening in developing countries: report of a WHO consultation," World Health Organization, 2002.

[27] E. M. Burd, "Human papillomavirus and cervical cancer," *Clinical Microbiology Reviews,* vol. 16, no. 1, pp. 1-17, 2003.

[28] M. Arbyn, M. Simon, E. Peeters, L. Xu, C. Meijer, J. Berkhof, K. Cuschieri, J. Bonde, A. Ostrbenk Vanlencak, F. Zhao, R. Rezhake, M. Gultekin, J. Dillner, S. de Sanjosé, K. Canfell, P. Hillemanns, M. Almonte, N. Wentzensen and M. Poljak, "2020 list of

human papillomavirus assays suitable for primary cervical cancer screening," *Clinical Microbiology and Infection,* vol. 27, no. 8, pp. 1083-1095, 2021.

[29] R. D. Burk, A. Harari and Z. Chen, "Human papillomavirus genome variants," *Virology,* vol. 445, no. 1-2, p. 232–243, 2013.

[30] L. Bruni, M. Diaz, X. Castellsagué, E. Ferrer, F. Bosch and S. de Sanjosé, "Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings," *The Journal of Infectious Diseases,* vol. 202, no. 12, pp. 1789-1799, 2010.

[31] Netherlands Comprehensive Cancer Organisation (IKNL), "National monitoring of the cervical cancer screening programme in the Netherlands 2021," Netherlands, 2022.

[32] The Swedish National Cervical Screening Registry (NKCx), "Förebyggande av livmoderhalscancer i Sverige. Verksamhetsberättelse och Årsrapport 2022 med data till och med 2021 [Prevention of cervical cancer in Sweden. Annual report with data up to and including 2021]," 2022.

[33] P. Veerus, R. Hallik, J. Jänes, K. Jõers, K. Paapsi, K. Laidra and K. Innos, "Human papillomavirus self-sampling for long-term non-attenders in cervical cancer screening: A randomised feasibility study in Estonia," *Journal of medical screening,* vol. 29, no. 1, pp. 53-60, 2021.

[34] Global Health Observatory data repository, "Cervical cancer screening. Response by country," World Health Organisation, 6 May 2022. [Online]. Available: https://apps.who.int/gho/data/view.main.UHCCERVICALCANCERv. [Accessed 9 February 2023].

[35] World Health Organization, "WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention, second edition," 2 ed., Geneva, 2021.

[36] European Commission, "Proposal for a Council Recommendation on strengthening prevention through early detection: A new EU approach on cancer screening replacing Council Recommendation 2003/878/EC," Office for Official Publications of the European Communities., Brussels, 2022.

[37] National Institute for Health Development; Estonian Health Insurance Fund, "HPV kodutest emakakaelavähi sõeluuringu täiendava meetodina: randomiseeritud pilootuuring [HPV self-sampling as an additional cervical cancer screenig method: a randomized pilot study]," 2022. [Online]. Available: https://tai.ee/et/valjaanded/hpv-kodutest-emakakaelavahi-soeluuringu-taiendava-meetodina-randomiseeritud-pilootuuring. [Accessed 9 February 2023].

[38] National Institute for Health Development, "HPV kodutesti juurutamine riiklikus emakakaelavähi sõeluuringus [Implementation of HPV self-sampling in the national organised cervical cancer screening programme]," 2022. [Online]. Available: https://www.tai.ee/et/teadustoo/teadustoo-ulevaade/naiste-tervise-uuringud/hpv-kodutesti-juurutamine-riiklikus. [Accessed 9 February 2023].

[39] National Institute for Health Development, "Üha enam naisi valib emakakaelavähi sõeluuringus osalemiseks kodutesti [More and more women are opting for a self-sampled test to participate in cervical cancer screening]," 5 May 2023. [Online]. Available: https://tai.ee/et/uudised/uha-enam-naisi-valib-emakakaelavahi-soeluuringus-osalemiseks-kodutesti. [Accessed 6 May 2023].

[40] S. DeWeerdt, "A global drive to eliminate cervical cancer," Nature Index, 22 April 2020. [Online]. Available: https://www.nature.com/articles/d41586-020-01036-x. [Accessed 17 April 2023].

[41] World Health Organization, Screening programmes: a short guide. Increase effectiveness, maximize benefits and minimize harm, Copenhagen: WHO Regional Office for Europe, 2020.

[42] Estonian Health Insurance Fund, "EHIF contract partners in prevention in health care," 23 February 2023. [Online]. Available: https://www.haigekassa.ee/inimesele/arsti-ja-oendusabi/haigekassa-lepingupartnerid/haiguste-ennetus. [Accessed 25 February 2023].

[43] K. Sild, "List of laboratory analyses," Tervise ja Heaolu Infosüsteemide Keskus, 2022. [Online]. Available: https://elhr.digilugu.ee/data/laboridList.html. [Accessed 6 March 2023].

[44] J. M. Kramer, Designing for Health Accessibility: Case Studies of Human-Centered Design, Berkeley: University of California, 2020.

[45] R. Nayar and D. C. Wilbur, "The Bethesda system for reporting cervical cytology: definitions, criteria, and explanatory notes," Springer, 2015.

[46] P. M. A, "The Bethesda System for reporting cervical cytology," *CytoJournal*, vol. 19, p. 28, 2022.

[47] National Institute for Health Development, "Estonian Cancer Screening Registry," 22 October 2019. [Online]. Available: https://en.tai.ee/en/r-and-d/registers/estonian-cancer-screening-registry. [Accessed 5 March 2023].

[48] Public Health Act, "§ 147. Estonian Cancer Screening Registry," 16 April 2021. [Online]. Available: https://www.riigiteataja.ee/en/eli/ee/529082019007/consolide/current. [Accessed 5 March 2023].

[49] Vähi sõeluuringute registri põhimäärus [Estonian Cancer Screening Registry Statute], 7 January 2022. [Online]. Available: https://www.riigiteataja.ee/akt/104012022004. [Accessed 5 March 2023].

[50] Eesti Teadusagentuur, "Vähi sõeluuringute register [Estonian Cancer Screening Registry]," 30 June 2015. [Online]. Available: https://www.etag.ee/rahastamine/programmid/loppenud-programmid/terve/programmi-tegevused/t3/skriining/. [Accessed 9 March 2023].

[51] E. Anion, Analysis of selected quality indicators in the Estonian cervical cancer screening program in 2016. Master Thesis., Tallinn: Tallinn University of Technology, 2016.

[52] National Audit Office of Estonia, Detection of malignant tumors and referring the patient to treatment, Tallinn, 2021.

[53] Health Statistics and Health Research Database, "Release Calendar 2023," 23 February 2023. [Online]. Available: https://statistika.tai.ee/index_en.html. [Accessed 5 March 2023].

[54] "Vähitõrje tegevuskava 2021–2030 [Cancer Control Plan for 2021-2030]," Sotsiaalministeerium, Tervise Arengu Instituut, Tallinn, 2021.

[55] Health Statistics and Health Research Database, "VSR17: Detected cancer cases in cervical cancer screening programme," 20 April 2022. [Online]. Available: https://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__02Haigestumus__07Soeluuringud/VSR17.px/table/tableViewLayout2/. [Accessed 19 April 2023].

[56] T. Lasn, "Vähi sõeluuringute dokumenteerimise nõuded. Rakendusjuhend arendajatele. [Cancer screening documentation requirements. Application guide for developers.]," Eesti e-tervise sihtasutus, 30 October 2015. [Online]. Available: https://pub.e-tervis.ee/manuals/V%C3%A4hi%20s%C3%B5eluuringu%20dokumenteerimise%20juhend. [Accessed 6 March 2023].

[57] Health and Welfare Information Systems Centre, "Ambulatoorse epikriisi täitmise juhend [Instructions for filling out an outpatient case summary]," 2021. [Online]. Available: http://pub.e-tervis.ee/manuals/Ambulatoorse%20epikriisi%20t%C3%A4itmise%20juhend. [Accessed 6 March 2023].

[58] Health and Welfare Information Systems Centre, " Saatekirja vastuse täitmise juhend [Instructions for filling out a reply to the reference letter]," 2022. [Online]. Available: http://pub.e-tervis.ee/manuals/Saatekirja%20t%C3%A4itmise%20juhend. [Accessed 6 Marcg 2023].

[59] Estonian Society for Laboratory Medicine, "Kvalitatiivse uuringu vastus [Qualitative analysis result]," 1 December 2021. [Online]. Available: https://pub.e-tervis.ee/classifications/Kvalitatiivse%20uuringu%20vastus. [Accessed 9 March 2023].

[60] International Organization for Standardization (ISO), "Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems (ISO 9241-210:2019)," 2019.

[61] R. Buchanan, "Human dignity and human rights: Thoughts on the principles of human-centered design," *Design issues,* vol. 17, no. 3, pp. 35-39, 2001.

[62] J. Giacomin, "What is human centred design?," *The design journal,* vol. 17, no. 4, pp. 606-623, 2014.

[63] IDEO, "About IDEO," [Online]. Available: https://www.ideo.com/about. [Accessed 17 April 2023].

[64] E. Gonen, "Tim brown, change by design: how design thinking transforms organizations and inspires innovation," *Markets, Globalization & Development Review,* vol. 4, no. 2, 2009.

[65] K. Erwin and J. A. Krishnan, "Redesigning healthcare to fit with people," *BMJ,* vol. 354, 2016.

[66] Ministry of Social Affairs, "National Health Plan 2020-2030," Tallinn, 2021.

[67] Health Board, "Terviseamet: emakakaelavähki on võimalik vaktsiiniga ennetada [Health Board: cervical cancer can be prevented with vaccine]," 23 January 2023. [Online]. Available: https://ta.vaktsineeri.ee/et/uudised/terviseamet-emakakaelavahki-voimalik-vaktsiiniga-ennetada. [Accessed 28 April 2023].

[68] A. Anttila and G. Ronco, "Description of the national situation of cervical cancer screening in the member states of the European Union," *Eur J Cancer* , vol. 45, no. 15, p. 2685–2708, 2009.

[69] Nordscreen, "Test coverage," [Online]. Available: https://nordscreen.org/analysis/cervical-cancer-screening/test-coverage/. [Accessed 28 April 2023].

[70] Health Statistics and Health Research Database, "VSR12: Cervical cancer screening coverage by examination in target population by age," 31 August 2022. [Online]. Available: https://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__02Haigestumus__07Soeluuringud/VSR12.px. [Accessed 1 March 2023].

[71] M. Arbyn, A. Anttila, J. Jordan, G. Ronco, U. Schenck, N. Segnan, H. Wiener, A. Herbert and L. von Karsa, "European Guidelines for Quality Assurance in Cervical Cancer Screening. Second edition--summary document," *Annals of oncology : official journal of the European Society for Medical Oncology,* vol. 21, no. 3, p. 448–458, 2010.

[72] L. Koreinik, Factors influencing women’s participation in cervical cancer screening in, Tallinn: Tartu University, 2019.

[73] R. Aarnio, E. Östensson, M. Olovsson, I. Gustavsson and U. Gyllensten, "Cost-effectiveness analysis of repeated self-sampling for HPV testing in primary cervical screening: a randomized study," *BMC Cancer,* vol. 20, no. 1, p. 645, 2020.

[74] Regionala cancercentrum, "National take care program for cervixcancer prevention," 2022.

[75] D. Grimes and K. Schulz, "Descriptive studies: what they can and cannot do," *Lancet,* vol. 359, no. 9301, pp. 145-149, 2002.

[76] Health Statistics and Health Research Database, "VSR11: Cervical cancer screening target population and women invited to screening by health insurance, county and age," 26 January 2023. [Online]. Available: https://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__02Haigestumus__07Soeluuringud/VSR11.px/. [Accessed 9 March 2023].

[77] A. Oštrbenk Valenčak, A. Šterbenc, K. Seme and M. Poljak, "Alinity m HR HPV Assay Fulfills Criteria for Human Papillomavirus Test Requirements in Cervical Cancer Screening Settings," *Journal of Clinical Microbiology,* vol. 58, no. 1, 2019.

[78] H. Frayle, S. Gori, M. Rizzi, B. Graziani, E. Vian, P. Giorgi Rossi and A. Del Mistro, "HPV testing for cervical cancer screening: technical improvement of laboratory logistics and good clinical performance of the cobas 6800 in comparison to the 4800 system," *BMC Womens Health,* vol. 19, no. 1, p. 47, 2019.

[79] SYNLAB, "Inimese papilloomiviiruse (HPV) diagnostiline paneel [HPV diagnostic panel]," [Online]. Available: https://synlab.ee/arstile/infomaterjalid/diagnostika-juhendid/. [Accessed 7 May 2023].

[80] P. D. P. Act, "Personal Data Protection Act," 12 December 2018. [Online]. Available: https://www.riigiteataja.ee/en/eli/523012019001/consolide. [Accessed 6 March 2023].

[81] Personal Data Protection Act, "§ 6. Processing of personal data for needs of scientific and historical research and official statistics," 23 January 2019. [Online]. Available: https://www.riigiteataja.ee/en/eli/523012019001/consolide. [Accessed 6 March 2023].

[82] Personal Data Protection Act, "§ 14. Principles of processing personal data," 23 January 2019. [Online]. Available: https://www.riigiteataja.ee/en/eli/523012019001/consolide. [Accessed 6 March 2023].

[83] Health Statistics and Health Research Database, "VSR13: Cervical cancer screening coverage by examination in target population by county," 31 August 2022. [Online]. Available: https://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__02Haigestumus__07Soeluuringud/VSR13.px/. [Accessed 1 April 2023].

[84] Health Statistics and Health Research Database, "VSR15: Cervical cancer screening invitational coverage and coverage by examination in women invited to screening by health insurance and county," National Institute for Health Development, 31 August 2022. [Online]. Available: https://statistika.tai.ee/pxweb/en/Andmebaas/Andmebaas__02Haigestumus__07Soeluuringud/VSR15.px/. [Accessed 2023 April 24].

[85] D. García, A. Cid-Arregui, M. Schmitt, M. Castillo, I. Briceño and F. Aristizábal, "Highly Sensitive Detection and Genotyping of HPV by PCR Multiplex and Luminex Technology in a Cohort of Colombian Women with Abnormal Cytology," *The Open Virology Journal,* vol. 5, p. 70–79, 2011.

[86] ECIS - European Cancer Information System, "Long term estimates of cancer incidence and mortality, for all cancers," 2023. [Online]. Available: https://ecis.jrc.ec.europa.eu/. [Accessed 8 February 2023].

[87] Estonian Cancer Screening Registry, Kasutajajuhed [user manual], Tallinn: Tervise Arengu Instituut, 2022.

[88] R. Rätsep, "Küsitlus emakakaelavähi sõeluuringu teadlikkusest ja osalusest Baltimaades [Cervical cancer screening awareness and participation survery in the Baltic counrties]," Roche Diagnostics Baltics, 2023.

[89] O. Nõmm, P. Veerus and K. Innos, "Pap-smear and cervical cancer in Estonia: a population-based case-control study," *European Journal of Public Health,* vol. 31, no. Supplement_3, 2021.

[90] S. Ozaki, K. Kato, Y. Abe, H. Hara, H. Kubota, K. Kubushiro, E. Kawahara and M. Inoue, "Analytical performance of newly developed multiplex human papillomavirus

genotyping assay using Luminex xMAP™ technology (Mebgen™ HPV Kit)," *Journal of Virological Methods,* vol. 204, pp. 73-80, 2014.

[91] S. Goldie, J. Goldhaber-Fiebert and G. Garnett, "Chapter 18: Public health policy for cervical cancer prevention: the role of decision science, economic evaluation, and mathematical modeling," *Vaccine,* vol. 24, no. 3, pp. 155-163, 2006.

Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Dolores Eamets

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis “Prevalence of high-risk human papillomavirus genotypes and their association with cervical cytology in Estonia: a population-based study”, supervised by Kaire Innos and Kadi Lubi.
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

06.05.2023

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.

Appendix 2 - Dataset

From the ESCR the following personal data was acquired on the study population:

1) ID-code;
2) date of birth;
3) place of residence (county).

From the ESCR the following screening data (HPV and LBC test results) was acquired on the study population:

1) ID-code;
2) laboratory analysis code and name (LOINC);
3) alternative code and name of laboratory analysis (treatment procedure invoice code);
4) medical document data (type, number, OID code, version and approval date)
5) evaluator data (name and register code of the healthcare service provider)
6) sample material data (collection date, HPV genotype and sample container identifier);
7) the date of evaluation;
8) result and the interpretation of the result;
9) reference value;
10) the reason for rejecting the sample material;
11) sample material;
12) the name and code of the health care service entered on the treatment bills of the Estonian Health Insurance Fund (codes 66644, 66821 of the list of health care services);
13) the date of service provision entered on the medical bills of the Estonian Health Insurance Fund;
14) the name and register code of the health care provider who provided the health care service entered on the treatment bills of the Estonian Health Insurance Fund.