

Meeste ja naiste palgaerinevuste statistika

Gender pay gap statistics

METOODIKAKOGUMIK
METHODOLOGICAL REPORT

Meeste ja naiste
palgaerinevuste statistika
Gender pay gap statistics

METOODIKAKOGUMIK
METHODOLOGICAL REPORT

Kogumiku on koostanud projekti „Palgaerinevuste parem kättesaadavus“ raames Teele Järvpõld ja Maretta Lunev. Kogumiku koostajad tänavad Merle Paatsi, kes abistas sotsiaaluuringute ühendandmebaasi idee loomisel ja elluviimisel.

The publication has been compiled by Teele Järvpõld and Maretta Lunev in the framework of the project “Increased availability of gender pay gap statistics“. The authors of the publication would like to thank Merle Paats who provided her assistance in creating and implementing the idea of a common database of social surveys.

Toimetanud Kairit Pöder

Inglise keelde tõlkinud OÜ Triangular

Inglise keele toimetanud Kairi Kübarsepp

Küljendanud Uku Nurges

Edited by Kairit Pöder

Translation into English by OÜ Triangular

English edited by Kairi Kübarsepp

Layout by Uku Nurges

Kirjastanud Statistikaamet, Tatari 51, 10134 Tallinn

Published by Statistics Estonia, Tatari 51, 10134 Tallinn

November 2015

November 2015

ISBN 978-9985-74-583-0

Autoriõigus/Copyright: Statistikaamet, 2015

Väljaande andmete kasutamisel või tsiteerimisel palume viidata allikale.

When using or quoting the data included in this issue, please indicate the source.

SISUKORD

SISSEJUHATUS	4
PROJEKTI TUTVUSTUS.....	5
SOOLISE PALGALÕHE ARVUTAMINE.....	6
ÜHENDANDMEBAASI LOOMINE	7
Ühendandmebaasi loomise alus	9
Eesti tööjõu-uuringu ja Eesti sotsiaaluuringu andmete ühendamine	10
Tunnuste ümberkodeerimine.....	10
Teadmata väärtused	14
Ühendandmebaasi erandjuhtumid	14
Palgatunnuse kvaliteet isiku-uuringutes	16
Maksu- ja Tolliameti andmete lisamine ühendandmebaasi	16
Maksu- ja Tolliameti andmebaasi struktuur	16
Maksu- ja Tolliameti andmete lisamine tööjõu-uuringu ja sotsiaaluuringu ühendandmebaasi	17
Maksu- ja Tolliameti andmebaasi põhjal koostatud tunnused	18
Laiendustegurite (kaalude) arvutamine	20
Leibkonnakaalud	21
Isikukaalud	21
Puuduvate väärtuste imputeerimine	21
Palga imputeerimine Eesti tööjõu-uuringus	21
Palga imputeerimine Eesti sotsiaaluuringus.....	22
Töötatud tundide arvu imputeerimine	23
Tööaja vormi imputeerimine	24
Ühendandmebaasi loomise kokkuvõte	25
PALGALÕHE JA SEDA MÕJUTAVAD NÄITAJAD.....	26
ÜHENDANDMEBAASI ANDMETE VÕRDLOS TÖÖTASU STRUKTUURI UURINGU ANDMETEGA	29
PALGAERINEVUSTE STATISTIKA STRATEEGIA.....	33
Lisa 1	35
Lisa 2 (I).....	56
Lisa 2 (II).....	57
Lisa 3.....	58

TABLE OF CONTENTS

INTRODUCTION	59
PROJECT INTRODUCTION	60
CALCULATING THE GENDER PAY GAP	61
ESTABLISHING THE COMBINED DATABASE	62
<i>Basis for establishing the combined database</i>	64
<i>Combining the data from the Estonian Labour Force Survey and the Estonian Social Survey</i>	65
<i>Re-coding variables</i>	65
<i>Unknown values</i>	69
<i>Exceptional cases in the combined database</i>	69
<i>Quality of wage variable in individual surveys</i>	72
<i>Adding Tax and Customs Board data to the combined database</i>	72
<i>Structure of Tax and Customs Board database</i>	72
<i>Adding Tax and Customs Board data to the combined database of the ELFS and the ESS</i>	73
<i>Variables produced based on the Estonian Tax and Customs Board database</i>	74
<i>Calculation of expansion factors (weights)</i>	76
<i>Household weights</i>	76
<i>Personal weights</i>	77
<i>Imputation of unknown values</i>	77
<i>Imputation of wages in the ELFS</i>	77
<i>Imputation of wages in the ESS</i>	79
<i>Imputation of number of hours worked</i>	80
<i>Imputation of working time format</i>	81
<i>Summary of establishing the combined database</i>	82
PAY GAP AND INDICATORS THAT AFFECT IT	83
COMPARISON BETWEEN THE DATA OF THE COMBINED DATABASE AND THE DATA OF THE STRUCTURE OF EARNINGS SURVEY	86
STRATEGY OF PAY GAP STATISTICS.....	90
Annex 1	92
Annex 2 (I).....	117
Annex 2 (II).....	118
Annex 3	119

SISSEJUHATUS

Sooline võrdõiguslikkus on üks Euroopa Liidu (EL) aluspõhimõtteid^a. Põhimõte „võrdne palk võrdse töö eest“ on EL-i aluslepingutes kasutusel alates 1957. aastast ja see on kirjas ka direktiivis 2006/54/EÜ^b, mis käsitleb meeste ja naiste võrdset kohtlemist tööturul. Soolisel palgalõhel on mitu põhjust: ebavõrdne sooline jaotumine hariduses ja tööturul, töö- ja pereelu ühitamise raskused, iganenud personalipraktika jm. Soolise palgalõhe arvutamiseks kogutakse praegu andmeid töötasu struktuuri uuringuga, mida korraldatakse iga nelja aasta tagant. Vahepealsetel aastatel arvutatakse sooline palgalõhe küsimustiku nr 1410 „Palk, lisa oktoobrikuu kohta“ põhjal^c. Et andmeid kogutakse ettevõttepõhiselt, ei ole võimalik koguda töötajate kohta tausttunnuseid, mille kaudu palgalõhe dünaamikat süvitsi analüüsida. Et hinnata palgaerinevuste põhjuseid, on vaja regulaarselt koguda mitmesuguseid vajalikke tausttunnuseid, nt töötaja haridustase, omandatud eriala, ameti- ja tegevusala, tööaja vorm, perekonnaseis, (väike)laste olemasolu jne.

Käesoleva metoodikakogumiku esimeses peatükis tutvustatakse lühidalt projekti „Palgaerinevuste statistika parem kättesaadavus“ ja selle peamisi eesmärke. Teises peatükis antakse lühiülevaade soolise palgalõhe arvutamise metoodikast ja andmeallikatest, mille põhjal seda näitajat senini on arvutatud. Kolmandas peatükis käsitletakse ühendandmebaasi loomist. Neljandas peatükis antakse ülevaade lisanäitajatest, mida saab peale soolise palgalõhe näitaja veel avaldada. Viiendas peatükis võrreldakse seni palgalõhe arvutamise aluseks olevate andmeallikate ja uue, projekti raames loodud ühendandmebaasi metoodilisi erinevusi ning andmete kokkulangevust. Viimane ehk kuues peatükk keskendub palgaerinevuste statistika strateegiale. Kogumiku lõpus on antud ülevaade ka kõikidest ühendandmebaasis avaldatud tunnustest.

^a http://europa.eu/rapid/press-release_IP-14-190_et.htm

^b <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:204:0023:0036:et:PDF>

^c http://www.stat.ee/public/files/aruandevormid/2015/21108141015_Aruandevorm_PDF-failina_et.pdf?t=1430208353

PROJEKTI TUTVUSTUS

Praegu kogutakse soolise palgalõhe arvutamiseks andmeid töötasu struktuuri uuringuga, mis on kohustuslik kõikidele EL-i riikidele. Uuringut korraldatakse EL-is iga nelja aasta tagant. Ka Eestis ei tehta seda uuringut alates 2008. aastast enam igal aastal (2009. aastal uuringut ei olnud). Kuna aga selles uuringus kogutakse andmed ettevõtete kaudu, saadakse väga vähe tausttunnuseid palgasaaja enda kohta.

Samal ajal on olemas teisigi andmeallikaid, kust saadakse igal aastal infot töötasu kohta, nt isiku-uuringud ja registrid. Juba olemasolevate andmeallikate kasutuselevõtt vähendaks andmeesitajate halduskoormust olulisel määral. Peale selle võimaldavad isiku-uuringud analüüsida naiste ja meeste palgaerinevusi süvitsi, kuna hõlmavad mitmesuguseid tausttunnuseid (nt töötaja haridustase, omandatud eriala, ameti- ja tegevusala, tööaja vorm, perekonnaseis, (väike)laste olemasolu jne).

Projektis „Palgaerinevuste statistika parem kättesaadavus“ selgitati välja, millised oleksid soolise palgalõhe seiramiseks sobivaimad juba olemasolevad andmeallikad (uuringud, registrid). 2013. aasta detsembris toimunud töötoas otsustati projekti raames kokku kutsutud eksperdirühmaga tehtud analüüside põhjal, et sobivaim oleks isiku-uuringutest liita Eesti tööjõu-uuringu ja Eesti sotsiaaluuringu valimid (valimi suurendamise eesmärgil), linkides neile juurde täpsed palgaandmed Maksu- ja Tolliametist. Seejärel asuti analüüsima, kas palgalõhe arvutamine nende uuringute alusel loodud ühendandmebaasi põhjal oleks kokkuvõttes lihtsam, kiirem ja odavam.

Projekti eesmärk oli seega välja töötada uus soolise palgalõhe arvutamise meetoodika, mille põhjal teha ettepanekuid metodoloogilisteks parendusteks ja palgalõhe arvutamise aluseks olevate andmeallikate kasutamiseks, mis ühtlasi vähendaksid andmeesitajate (ettevõtete) ja Statistikaameti halduskoormust.

Uute andmeallikate kasutamine võimaldab arvutada ja avaldada soolist palgalõhet järgmiselt:

- ajakohaselt, mis tähendab, et eelmise aasta andmed on võimalik avaldada hiljemalt järgmise aasta esimesel poolel (nt 2015. aasta sooline palgalõhe avaldatakse 2016. aasta esimese poole jooksul);
- koos relevantsete näitajatega, mis kirjeldavad soolise palgalõhe tagamaid ja võimaldavad poliitika väljatöötajatel tugineda teadmuspõhisele analüüsile.

Projekti tarvis on kokku kutsutud eksperdirühm ja korraldatud regulaarselt töötubasid, kus tööruhma kuuluvad eksperdid ning huvirühmade esindajad on saanud kaasa rääkida nii andmeallikates, andmebaasi tunnustes, lisanäitajates kui ka meetoodilistes otsustes. Täpsem info töötubades arutletud teemade ja vastu võetud otsuste kohta on kättesaadav projekti veebilehel^a.

Projekti rahastatakse 2009.–2014. aasta Norra toetuste programmi „Sooline võrdõiguslikkus ning töö- ja pereelu tasakaal“ kaudu. Ajakohane info projekti tegevuste kohta on kättesaadav Statistikaameti veebilehel rubriigis „Majandus“ > „Palk ja tööjõukulu“.

^a <http://www.stat.ee/palgaerinevuste-statistika-parem-kattesaadavus>

SOOLISE PALGALÕHE ARVUTAMINE

Soolise palgalõhe arvutamist hakkas EL-i statistikaamet Eurostat ühtlustama 2005. aastal. Kuni 2005. aastani kasutati soolise palgalõhe arvutamiseks riigiti eri allikaid – administratiivsed andmed, tööjõu-uuring (ingl Labour Force Survey), sotsiaaluuring (ingl EU Statistics on Income and Living Conditions) või mõni muu riigis tehtav uuring. Selle tulemusena olid definitsioonides, üldkogumis, valimi suuruses jm erinevused, mis mõjutasid soolise palgalõhe näitaja kvaliteeti ja riikidevahelist võrreldavust. Seetõttu tekkis vajadus minna üle ühisele andmeallikale ja töötati välja ühtne soolise palgalõhe arvutamise meetodika.

EL-is kasutatakse soolise palgalõhe jälgimiseks Eurostati avaldatavat korrigeerimata palgalõhe näitajat (ingl *gender pay gap in unadjusted form*), mida avaldatakse igal aastal kõikide liikmesriikide kohta. Sooline palgalõhe näitab naiste ja meeste palgaerinevust. Näitaja arvutamisel kasutatakse brutotunnipalka ja see näitab, mitu protsenti on naiste brutotunnipalk meeste omast väiksem. Valem on järgmine:

$$\frac{\text{meeste keskmine brutotunnipalk} - \text{naiste keskmine brutotunnipalk}}{\text{meeste keskmine brutotunnipalk}} * 100$$

Eurostati avaldatud andmed soolise palgalõhe kohta kogutakse töötasu struktuuri uuringuga (ingl Structure of Earnings Survey), mida EL-i liikmesriigid peavad tegema iga nelja aasta tagant. Viimane töötasu struktuuri uuring toimus 2014. aastal. Kuna palgalõhe kohta avaldatakse andmeid igal aastal, teeb osa riike uuringut samuti igal aastal, teised kasutavad vahepealsetel aastatel muid andmeallikaid. Eestis kasutatakse vahepeal palgalõhe arvutamiseks küsimustiku „Palk, lisa oktoobrikuu kohta“ andmeid^a. Iga aasta (v.a sel aastal, kui toimub töötasu struktuuri uuring) oktoobrikuus täidavad valimisse sattunud ettevõtted küsimustikus mooduli, mille andmete põhjal arvutataksegi sooline palgalõhe.

^a <http://www.stat.ee/65596>

ÜHENDANDMEBAASI LOOMINE

Projekti „Palgaerinevuste statistika parem kättesaadavus“ üks peamisi eesmärke on luua ühendandmebaas, kus on peale palgaandmete ka muud olulised ja vajalikud tunnused, mille põhjal on võimalik palgalõhe komplekselt analüüsida ning mis omakorda aitavad mõista palgalõhe olemust ja selgitada selle tagamaid. Ühtse ja põhjaliku andmebaasi olemasolu annab ühelt poolt võimaluse teha süvaanalüüse, mis annavad suuna ja sisendi poliitikate väljatöötamiseks, ning teiselt poolt võimaluse jälgida ühiskonna arengut.

Ühendandmebaasi aluseks olevate võimalike andmeallikate väljaselgitamist ja arutelu andmebaasis avaldatavate tunnuste üle alustati kohe pärast projekti algust. 2013. aasta detsembris toimus projektiks kokku kutsutud töörühma spetsialistidega nõupidamine, kus võeti vastu peamised otsused selle kohta, milliste andmeallikate põhjal ühendandmebaas luuakse ja millised tunnused seal avaldatakse. Ka tarbijatele anti võimalus kaasa rääkida – Statistikaameti veebilehel oli võimalik esitada ettepanekuid. Töötoa materjalid on kättesaadavad projekti veebilehel^a.

Ühendandmebaasi loomisel tehtud peamised otsused:

- andmebaas luuakse Eesti tööjõu-uuringu ja Eesti sotsiaaluuringu põhjal;
- tööjõu-uuringu rotatsiooni tõttu osaleb üks isik aasta jooksul uuringus kõige rohkem kaks korda. Otsustati, et kõiki isikuid käsitletakse kui esimest korda uuringusse sattunud ehk rotatsiooni arvesse ei võeta;
- arvesse võetakse kõik isikud, kes on nende uuringute kohaselt palgatöötajad;
- Maksu- ja Tolliameti andmetest lingitakse ühendandmebaasile juurde palgaandmed ehk tööne tulu.

Ühtlasi lepiti kokku, millised tunnused peavad olema loodavas ühendandmebaasis, mille põhjal oleks võimalik teha edasisi põhjalikke analüüse ja arvutada teisi olulisi näitajaid, mida koos palgalõhega regulaarselt avaldada. Lõplikus ühendandmebaasis on olemas järgmised tunnused (vt täpsemalt lisa 1):

- sugu;
- sünnikuu;
- sünniaasta;
- vanus uuringuaasta alguse seisuga;
- rahvus;
- kodakondsus;
- seaduslik perekonnaseis;
- tegelik perekonnaseis;
- sünniriik;
- elukoht;
- leibkonna elukoht;
- leibkonna elukoht asustustiheduse järgi;
- kodune keel;
- isiku leibkonnas elavate laste arv;
- isiku leibkonnas elavate laste vanused;
- kõrgeim omandatud haridustase;
- omandatud eriala;
- vanus esimest korda pidevale tööle asumisel;
- viimasele töökohale asumise aasta;
- sotsiaal-majanduslik seisund;

^a <http://www.stat.ee/75328>

- ametiala;
- tegevusala;
- tööaja vorm;
- tavaline nädalas töötatud tundide arv;
- töökoha asukoht;
- töötajate arv ettevõttes;
- tööalane seisund;
- staatus (ettevõtja või palgatöötja);
- kõrvaltöö olemasolu;
- juhtimiskohustuste olemasolu;
- töölepingu tüüp;
- ametiühingusse kuuluvus;
- ettevõtte majandussektor;
- ettevõtte omandivorm;
- ettevõtte suurus;
- uuringus öeldud tavaline brutokuupalk;
- Maksu- ja Tolliameti andmete põhine töine tulu.

Peale andmebaasi loomise oli vaja välja selgitada, milliseid näitajaid koos palgalõhega regulaarselt arvutama ja avaldama hakata. Kõiki soolise palgalõhega seotud näitajaid üksnes ühendandmebaasi põhjal leida ei saagi, sest selleks vajalikke andmeid uuringutega ei koguta või puudutavad need ühiskonna protsessid niivõrd väikest inimeste hulka, et uuringuga ei olegi võimalik nende kohta andmeid koguda. Seepärast otsustati järgmisena, millised on need näitajad, mis aitavad mõista soolise palgalõhe tekke ja püsimise põhjusi ning mida koos palgalõhega kindlasti avaldada tuleks.

2013. aasta detsembris toimunud töötoas otsustati, et näitajate aluseks võetakse Sotsiaalministeeriumi kogumikus „Sooline palgalõhe Eestis. Artiklite kogumik“ avaldatud soovituslikud näitajad, mille abil saab jälgida töö tasustamise erinevusi ja (struktuursete) tegurite rolli selles^a.

2015. aasta märtsis toimunud ekspertide töörühma kohtumisel otsustati koos palgalõhe aegreaga (2006–2013) avaldada järgmised näitajad:

- keskmine sooline palgalõhe;
- sooline palgalõhe järgmiste tunnuste järgi:
 - vanus
 - rahvus
 - elukoht
 - haridustase
 - kõrgeima omandatud hariduse eriala
 - ametiala
 - tegevusala
 - tööaja vorm
 - juhtimiskohustuste olemasolu
 - tööandja liik
 - tööandja suurus
 - perekonnaseis
 - laste olemasolu

^a https://www.sm.ee/sites/default/files/content-editors/Ministeerium_kontaktid/Valjaanded/toimetised_20112.pdf

- tegevusala- ja ametialapõhine sooline segregatsioon tööturul;
- meeste ja naiste osatähtsus juhtivtöötajate hulgas;
- meeste ja naiste osatähtsus ettevõtjate hulgas;
- meeste ja naiste loodud ettevõtete osatähtsus alustavate ettevõtete hulgas;
- meeste ja naiste osatähtsus registreeritud FIE-de hulgas;
- osalise tööajaga töötajate osatähtsus nais- ja meessoost palgatöötajate hulgas;
- lapsehooldusteenuste kättesaamatusest tingitud passiivsus
- meeste osatähtsus vanemahüvitise saajate hulgas;
- naiste ja meeste keskmine vanemahüvitise summa kuus;
- meeste osatähtsus lapsehoolduspuhkuse kasutajate hulgas;
- isapuhkuse kasutamine;
- ajakasutus soo järgi;
- lapsevanemaks olemise mõju hõivelõhele.

Regulaarselt ei avaldata selgitatud ja selgitamata palgalõhe näitajaid (aastate 2011 ja 2012 kohta; järgmine soovituslik 2015. aasta kohta).

Ekspertide rühma koostamise materjalid ja lisainfo vastu võetud otsuste kohta on kättesaadavad projekti veebilehel^a.

Ühendandmebaasi loomise alus

Ühendandmebaasi kokkupanek algab Eesti tööjõu-uuringu (ETU) ja Eesti sotsiaaluuringu (ESU) tunnuste ühendamise. Esmalt analüüsitakse tunnuste kogumise meetodikat kummaski uuringus ja erinevuste korral kodeeritakse originaaltunnused ümber selliselt, et tunnused oleksid omavahel võrreldavad. Seejärel on võimalik uuringute andmebaasid omavahel ühendada.

Statistikaametis loodud unikaalse identifitseeriva tunnuse alusel liidetakse ühendandmebaasile Maksu- ja Tolliameti (MTA) maksukohustuslaste registri andmed. Siinkohal aga tuleb arvestada, et MTA andmebaasis ei vasta igale isikule üks kirje, vaid üks kirje tähendab ühele isikule tehtud väljamakseid ühelt ettevõttelt ühes kuus. Seega selleks, et MTA andmed isiku-uuringutes osalenud isikute omadele liita, tuleb MTA andmebaas kõigepealt isiku tasandile ümber korraldada nii, et iga isiku kohta oleks ainult üks kirje.

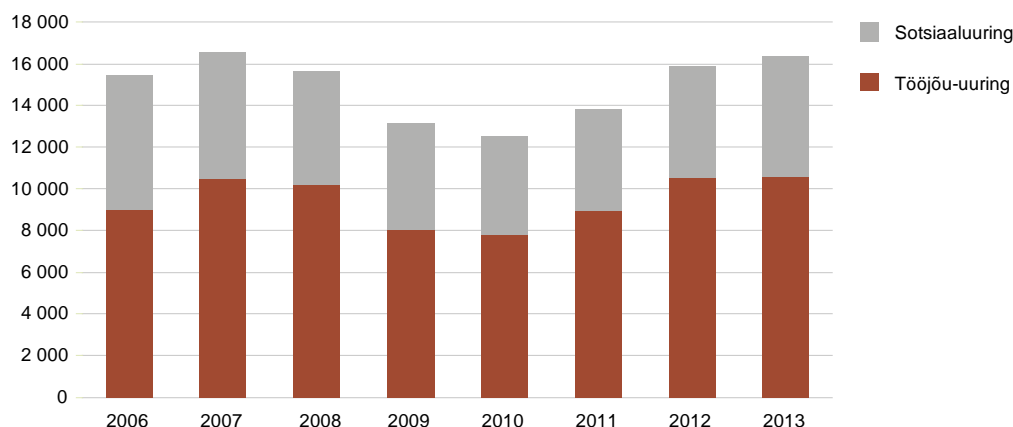
Ühendandmebaasile leitakse uus laiendustegur (kaal), mis teeb tulemused üldistatavaks kogu rahvastikule.

Et ETU aastavalim on suurem kui ESU-l, on ühendandmebaasis ETU-s osalenud inimesi rohkem kui ESU-s osalenuid. Joonisel 1 on kummagi uuringu isikute arv aastatel 2006–2013. Eraldi on välja toodud ka palgatöötajate arv, sest palgatöötajate põhjal arvutatakse sooline palgalõhe.

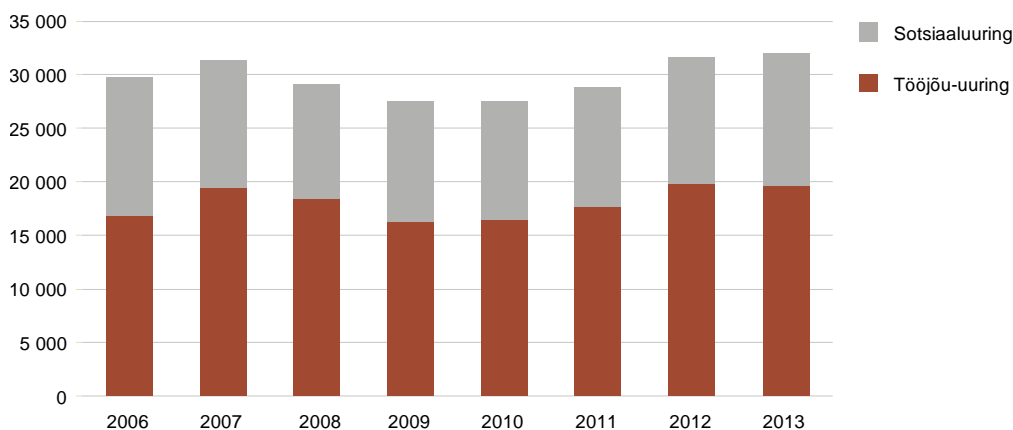
^a <http://www.stat.ee/183300>

Joonis 1. Isikute ja palgatöötajate arv ühendandmebaasis uuringu järgi, 2006–2013

Palgatöötajad



Isikud



Järgnevalt kirjeldatakse ühendandmebaasi kokupanekut, tekkinud probleeme ja nende lahendusi ning eri andmeallikatel põhinevate tunnuste ühilduvust.

Eesti tööjõu-uuringu ja Eesti sotsiaaluuringu andmete ühendamine

Tunnuste ümberkodeerimine

Kahe uuringu tunnuste ühendamise eelduseks on, et tunnuseid kogutakse mõlemas uuringus ja sealjuures ühtemoodi, mis tähendab, et küsimus on sama tähendusega ning vastusevariandid on samad. Ühendandmebaasi loomisel tekib probleem juhul, kui mõnda tunnust kogutakse vaid ühes uuringus, tunnust kogutakse uuringutes erinevalt (nt vastusevariandid erinevad) või küsitakse tunnust eri rühmadelt. Sellisel juhul kodeeritakse tunnus ümber.

Ainult ühe uuringuga kogutavad tunnused

ETU ja ESU tunnuste ühendamisel tekib probleem töökoha asukohaga, mida küsitakse ainult ETU-s. ESU-s osalenutel on see tunnus puudu.

Uuringutes erinevalt kogutavad tunnused

Teine probleem on, et tunnuseid kogutakse ETU-s ja ESU-s erinevalt.

Esimene niisugune tunnus on tööaja vorm, s.t kas isik töötab osalise või täistööajaga. See info peaks olema olemas kõikide nende isikute kohta, kes uuringu toimumise ajal töötavad. ETU-s hindab vastaja

subjektiivselt, kas ta töötab täis- või osalise tööajaga. ESU-s küsitakse üksnes hinnangut oma sotsiaal-majanduslikule staatusle, sh tööaja vormile. Valikuvariandid on järgmised:

- täistööajaga palgatöötaja;
- osalise tööajaga palgatöötaja;
- täistööajaga ettevõtja-tööandja, üksikettevõtja, palgatöötajata talupidaja, vabakutseline, palgata töötaja pereettevõttes (talus);
- osalise tööajaga ettevõtja-tööandja, üksikettevõtja, palgatöötajata talupidaja, vabakutseline, palgata töötaja pereettevõttes (talus);
- töötu;
- (üli)õpilane, tasustamata praktikant;
- eel- või vanaduspensionär;
- töövõimetuspensionär;
- kodune, lapsehoolduspuhkusel;
- ajateenija;
- muu mitteaktiivne.

Et mõlemas uuringus kogutud tunnus tugineks vastaja enda hinnangule, tuleb ESU-s kasutada samuti subjektiivset hinnangut. Probleem on aga nende isikutega, kellel on kaks seisundit, nt kui palgatöötaja on samal ajal ka pensionär. Need isikud ei pruugi enda peamiseks seisundiks pidada töötamist, mistõttu jääb teadmata, kas isik töötas tegelikult osalise või täistööajaga. Selle tõttu jääb ESU-s esialgu tööaja vorm mõnel juhul teadmata, kuid hilisema töötamise käigus see imputeeritakse.

Mõlemas uuringus küsitakse aga tavalist nädalas töötatud tundide arvu, mille põhjal saab samuti hinnata, kas isik töötas osalise või täistööajaga. Selleks tuleb kõigepealt vaadata, kui hästi läheb kokku vastaja hinnang ja tavaliselt töötatud tundide arvu põhine tööaja vorm. Rahvusvahelise määratluse järgi on osalise ajaga töötajad need, kes töötavad nädalas alla 35 tunni. Vähemalt 35 tundi nädalas töötavad isikud on täistööajaga töötajad.

Kahe tunnuse võrdlus näitab, et üle 90% isikute puhul lähevad inimese hinnang ja tavaline nädalas töötatud tundide arv kokku. See tähendab, et nende isikute puhul, kellel enesehinnanguline tööaja vormi tunnus puudub, saab selle tuletada tavalisest nädalas töötatud tundide arvust. Täpsem analüüs tööaja vormi tunnuse loomise kohta on kättesaadav kasutatavate andmeallikate ühilduvusanalüüsis^a.

Tabelis 1 on kogu ETU ja ESU ühendandmebaasis olevate kirjete ja töötavate isikute arv. Ühtlasi on tabelis märgitud nii nende isikute arv, kellele oli võimalik määrata tööaja vorm, kui ka nende oma, kellele seda ei olnud võimalik määrata. Sellisel juhul leiti netopalk statistiliste meetoditega teiste tunnuste kaudu ehk netopalk imputeeriti. Imputeerimise täpsemat meetodikat on kirjeldatud peatükis „Tööaja vormi imputeerimine“.

Tabel 1. Tööaja vormi tunnuse olemasolu ühendandmebaasis, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Kirjed	29 793	31 284	29 131	27 466	27 604	28 734	31 616	31 971
Töötavad isikud	16 803	18 092	16 933	14 340	13 690	15 161	17 443	17 915
Tööaja vorm olemas	16 803	18 090	16 932	14 339	13 686	15 155	17 434	17 901
Tööaja vorm puudu	0	2	1	1	4	6	9	14

^a <http://www.stat.ee/dokumendid/78174>

ETU-s ja ESU-s kogutakse erinevalt ka tavalist nädalas töötatud tundide arvu, mille tõttu ei ole võimalik neid tunnuseid automaatselt ühendada. Küsimus on mõlemas uuringus ühesugune: „Mitu tundi nädalas teete seda tööd tavaliselt?“.

ETU-s on võimalik anda ka järgmine vastus: „Raske öelda, tundide arv on nädalate või kuude lõikes väga erinev“. Samuti on ETU-s küsimus uuringunädalal töötatud tundide arvu kohta (päeva kaupa), mida ESU-s ei ole.

Tavaline nädalas töötatud tundide arv on palgaerinevuste statistika arvutamisel äärmiselt oluline tunnus, sest just selle abil leitakse tunnipalk, mis on naiste ja meeste palgavõrdluse alus. Et teadmata väärtuste osatähtsust minimeerida, otsustati nende isikute puhul, kes valisid ETU-s tavalise nädalas töötatud tundide arvu kohta vastuseks variandi „Raske öelda, tundide arv on nädalate või kuude lõikes väga erinev“, kasutada uuringunädalal töötatud tundide arvu. Põhjalikum tunnuse analüüs on kättesaadav kasutatavate andmeallikate ühilduvusanalüüsis^a.

Tabelis 2 on ühendandmebaasis olevate kirjete ja töötavate isikute arv. Ühtlasi on märgitud tabelis nii nende isikute arv, kellele oli võimalik määrata nende tavaline nädalas töötatud tundide arv, kui ka nende arv, kellele seda ei olnud võimalik määrata.

Tabel 2. Tavalise nädalas töötatud tundide arvu tunnuse olemasolu ühendandmebaasis, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Kirjed	29 793	31 284	29 131	27 466	27 604	28 734	31 616	31 971
Töötavad isikud	16 803	18 092	16 933	14 340	13 690	15 161	17 443	17 915
Tavaline nädalas töötatud tundide arv olemas	16 799	18 087	16 919	14 334	13 665	15 030	17 250	17 727
Tavaline nädalas töötatud tundide arv puudu	4	5	14	6	25	131	193	188

Probleeme on aga ettevõtte suuruse tunnusega. Töötajate arvu ettevõttes kogutakse ETU-s ja ESU-s erinevalt. Mõlemas uuringus on küll küsimus „Kui palju on selles ettevõttes/asutuses/üksuses töötajaid?“, kuid vastusevariandid on erinevad. Samal ajal on mõlemas uuringus võimalik vastused ümber kodeerida nii, et ühendatud andmebaasis oleks üks tunnus. Tabelis 3 on näidatud, kuidas on ETU-s ja ESU-s olev töötajate arvu tunnus ühendandmebaasi jaoks ümber kodeeritud.

Tabel 3. Ettevõtte töötajate arvu tunnuse ümberkodeerimine

Ühendandmebaasis	Töötajate-uuringus	Sotsiaaluuringus
1. 1 kuni 10	1. 1 kuni 10	1. 1 kuni 10 5. Ei tea täpselt, kuid vähem kui 11
2. 11 kuni 19	2. 11 kuni 19	2. 11 kuni 19
3. 20 kuni 49	3. 20 kuni 49	3. 20 kuni 49
4. 50 või rohkem	4. 50 kuni 99 5. 100 kuni 199 6. 200 kuni 499 7. 500 kuni 999 8. 1000 või rohkem	4. 50 või rohkem
5. Teadmata		6. Ei tea täpselt, kuid rohkem kui 10

Kuna ESU-s on töötajate arvu küsimusel vastusevariant „Ei tea täpselt, kuid rohkem kui 10“, jääb selles uuringus ettevõtte suurus teadmata sagedamini kui ETU-s (vt tabel 4).

^a <http://www.stat.ee/dokumendid/78174>

Tabel 4. Ettevõtte töötajate arvu tunnuse olemasolu andmebaasides uuringu järgi, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Tööjõu-uuring								
Töötavad isikud kokku	9 834	11 547	11 058	8 791	8 562	9 819	11 584	11 628
Töötajate arv ettevõttes teadmata	454	545	94	94	113	106	144	147
Sotsiaaluuring								
Töötavad isikud kokku	6 969	6 545	5 875	5 549	5 128	5 342	5 859	6 287
Töötajate arv ettevõttes teadmata	423	448	283	238	198	258	258	234

Erinevalt kogutakse ETU-s ja ESU-s ka andmeid palgatöötaja kuupalga kohta. ESU-s küsitakse vastajalt, kui suur on tema keskmine netokuupalk sellel töökohal. Kui vastaja ei taha või ei oska sellele küsimusele vastata, on küsitlejal võimalik märkida „Ei tea“ või „Keeldub vastamast“.

ETU-s küsitakse kõigepealt vastaja viimase kuu neto- või brutotasu tema põhitöökohal. Seejärel küsitakse, kas viimase kuu töötasu oli sama suur kui tavaline kuupalk. Kui viimase kuu töötasu tavalisest erineb, küsitakse vastaja käest tema tavalist neto- või brutokuupalka põhitöökohal. Mõlema palgaküsimuse juures on küsitlejal võimalik märkida „Ei tea“ või „Keeldub vastamast“, kui vastaja ei taha või ei oska küsimusele vastata. Kui vastaja ei soovi vastata viimase kuu tasu kohta, küsitakse tema viimase kuu netokuupalga vahemikke.

Et ESU-s ei ole küsimust viimasel kuul saadud palga kohta, kasutatakse ühendandmebaasis tavalise kuupalga tunnust. Seega kodeeritakse ETU-s küsitud tunnused ümber järgmiselt:

- juhul kui viimase kuu töötasu tavalisest erines, kasutatakse tavalise kuupalga küsimust;
- juhul kui viimase kuu töötasu oli sama mis tavaliselt, kasutatakse viimase kuu neto- või brutopalka põhitöökohal;
- juhul kui vastati bruto-, mitte netopalka kohta, teisendatakse brutopalk ümber netopalgaks, võttes arvesse ainult tulumaksu.

Keskmise netokuupalga väli jääb tühjaks nendel, kes ei soovi või ei oska palgaküsimusele vastata. Sel juhul leitakse nende netopalk statistiliste meetoditega teiste tunnuste kaudu, s.t netopalk imputeeritakse. Imputeerimise meetodikat on käsitletud peatükkides „Palga imputeerimine ETU-s“ ja „Palga imputeerimine ESU-s“.

Eri rühmadelt kogutud tunnused

Osa tunnuseid kogutakse ETU-s ja ESU-s eri rühmadelt. ETU-s küsitakse juhtimiskohustuste olemasolu ning ettevõtte sektori ja omandivormi kohta kõikidelt töötavatel isikutelt, ESU-s aga üksnes palgatöötajatelt. Seetõttu jäetakse nende tunnuste puhul kummaski uuringus alles üksnes palgatöötajate vastused. Samuti küsitakse ETU-s ettevõtte suurust kõigilt töötavatel isikutelt, aga 2011. aasta ESU-s ja varasemates ei küsitud ettevõtte suurust üksikettevõtjatelt ega palgata töötajatelt pereettevõttes/talus. Seega otsustati ühendandmebaasis 2011. aasta ESU-s ja varasemates üksikettevõtjatele määrata töötajate arvu väärtuseks „1“ ning pereettevõttes palgata töötajate puhul „Teadmata“.

Tunnused ühendandmebaasis

Lõplikus ühendandmebaasi failis on kokku 61 tunnust, mis tuginevad ETU-s ja ESU-s kogutud andmetele. Isiku-uuringutega kogutavate tunnuste põhjalik loetelu on kogumiku lõpus olevas lisas (vt lisa 1).

Teadmata väärtused

Isiku-uuringutega kogutud andmete puhul ei õnnestu alati kõiki andmeid vastaja käest saada, mistõttu jääb osa tunnuseid teadmata (kui vastaja ei soovi või ei oska vastust anda). Kui vastaja lõplikuks vastuseks jääb „Ei oska öelda“ või „Keeldun vastamast“, käsitletakse neid hilisemas analüüsis teadmata väärtustena.

Tausttunnused, nagu vastaja sugu, vanus, rahvus, perekonnaseis, elukoht jms, peaksid olema kõigi vastanud isikute kohta olemas, kuid paljude teiste tunnuste puhul esineb ka teadmata väärtuseid. Ühendandmebaasis on mõne tunnuse puhul teadmata väärtuste hulk andmemetoodiliste tööde tõttu veelgi suurem: et kahe uuringu tunnuseid oleks võimalik käsitleda võrdsetena, tuleb mõnikord tunnused ümber kodeerida ning seetõttu võib juhtuda, et osa väärtuseid tuleb ümber kodeerida teadmata väärtusteks. Detailne analüüs teadmata väärtustest ühendandmebaasis on kättesaadav kasutatavate andmeallikate ühilduvusanalüüsis^a.

Ühendandmebaasi erandjuhtumid

Kuna ETU ja ESU ankeedid ning küsimuste esitamise viisid ei ole alati täpselt ühesugused, oli ühendandmebaasi kokkupanekul erandjuhtumeid, mida tuli kahe uuringu andmebaaside kokkupanemisel arvestada.

Esimene läbiv küsimus puudutab ühendandmebaasi vanuserühmi. ETU-s vastavad isikuküsitlusele kõik 15–74-aastased, ESU-s aga isikud alates 15. eluaastast, kuid andmebaasid, mida kasutati ka ühendandmebaasi kokkupanekul, on kohandatud Eurostati vajadustele, s.t nendes on isikud alates 16. eluaastast.

Teine läbiv nüanss, mida tuleb ESU andmete puhul arvestada, on imputeeritud sissetulekutunnused. Alates 2007. aastast on ESU-s sissetulekutunnus imputeeritud kõikidele nendele, kes oleksid pidanud vastama isikuküsitlusele, kuid ei ole seda teinud (keeldumine või kadu). Nende sissetulekutunnuste imputeerimine ESU-s on vajalik selleks, et arvutada kogu leibkonna sissetulek. Kuna ühendandmebaasi loomisel on uuringutest võetavate andmete puhul rõhuasetus eelkõige hariduse ja töötamisega seotud tausttunnustel, on need nn asendusankeedid välja jäetud. Samuti ei ole neid arvestatud ühendandmebaasi kaalude arvutamisel.

Tabel 5. Lõplik kirjete arv ühendandmebaasis uuringu järgi, 2006–2013

Aasta	Sotsiaaluuring	Tööjõu-uuring
2006	13 007	16 786
2007	11 875	19 409
2008	10 761	18 370
2009	11 220	16 246
2010	11 114	16 490
2011	11 076	17 658
2012	11 775	19 841
2013	12 348	19 623

Järgnevalt on esitatud ühendandmebaasi loomisel esinenud erandjuhtumid.

- Kuigi uuringutes küsitakse vastaja isikukoodi, leidub siiski ankeete, kus see on jäetud vastamata ja kus pole seda õnnestunud ka töötuse käigus tagantjärele lisada. Kuna isikukoodide täpsustamiseks on tarvis teada isiku nime ja sünniaega, ei õnnestu see täpsete andmete puudumisel. Isikukoodi puudumisega olid suurimad probleemid 2006. aasta ETU andmebaasis, kus isikukood oli puudu 4426 isikul (vt tabel 6).

^a <http://www.stat.ee/dokumendid/78174>

Tabel 6. Isikukood uuringu järgi, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Isikukood olemas	25 345	30 835	28 389	27 152	27 275	28 448	31 423	31 821
Isikukood puudu tööjõu-uuringus	4 426	373	701	252	288	262	168	129
Isikukood puudu sotsiaaluuringus	22	76	41	62	38	24	25	21

- Ühendandmebaasi aluseks olevat palgaküsimust 2006. aasta ESU-s veel ei olnud, mistõttu selle aasta kohta on uuringuga kogutud palgaandmed puudu. Isikuankeedis olevat küsimust netokuupalga kohta („Kui suur on Teie keskmine netokuupalk sellel töökohal?“) on küsitud alates 2007. aastast.
- Aastatel 2006–2008 oli ETU-s probleeme elukohatunnusega, sest elukoht on isikupõhises andmebaasis olemas ainult maakonna tasemel, sh on Tallinn Harjumaa sees. Niisugune andmeesis ei võimaldanud aga Tallinna eraldi välja tuua ega leida leibkonna elukohta asustustiheduse järgi, milleks on vaja kolmekohalist valla koodi. Probleemi lahendamiseks võeti kõigile isikuküsitlusele vastanutele nende elukoht vastava aasta algfaasist ja Tallinn kodeeriti eraldi. Kuna ühendandmebaasis on vaja leibkonna, mitte isiku elukohta, määrati leibkonna elukohaks leibkonna vastutava liikme elukoht. Juhul kui vastutav leibkonnaliige elas välismaal, määrati leibkonna elukoht teiste leibkonnaliikmete põhjal.
- Aastatel 2006–2008 ei olnud ESU-s küsimust koduse keele kohta, mille tõttu on nende aastate ühendandmebaasides koduse keele tunnus olemas vaid nendel, kes osalesid ETU-s.
- ESU-s ei küsitud 2006. aastal ettevõtte/asutuse omandivormi kohta, seega on puudu ka avaliku ja erasektori jaotus. Sellel aastal puuduvad ka andmed isiku töölepingu tüübi kohta ESU-s osalenutel ning ühendandmebaasis on need tunnused olemas vaid neil, kes osalesid ETU-s.
- Kuna aastatel 2006 ja 2007 ei küsitud ESU-s, kas isik elab koos oma seadusliku abikaasaga, tekkis ühendandmebaasis probleeme isiku tegeliku perekonnaseisu kindlaksmääramisega. Seega kasutati selle tunnuse arvutamisel leibkonnasuhete maatriksi andmeid.
- Aastatel 2006 ja 2007 oli haridust puudutav küsimus isiku-uuringutes veidi teistsugune kui hilisematel aastatel, millest võivad olla põhjustatud mõningased muutused ISCED97 tasemetes 3b, 3a ja 4b üleminekul 2008. aastale (vt tabel 7).

Tabel 7. Uuringus osalenute haridustase ISCED97 klassifikaatori järgi, 2007, 2008

(protsenti)

	2007		2008	
	ETU	ESU	ETU	ESU
Tase 0	0,5	0,5	0,4	0,4
Tase 1	4,8	5,0	4,4	4,8
Tase 2c	0,1	0,5	0,1	0,3
Tase 2a	20,9	22,4	21,1	22,4
Tase 3c	4,9	5,7	4,1	5,0
Tase 3b	7,7	8,3	3,1	4,9
Tase 3a	32,6	31,2	41,4	38,3
Tase 4b	5,0	6,5	1,0	2,6
Tase 5b	9,6	7,8	8,7	8,1
Tase 5a	13,8	11,7	15,6	12,9
Tase 6	0,2	0,3	0,1	0,3
Teadmata	0,0	0,0	0,0	0,0

- Ettevõtte töötajate arvu küsimuses oli aastatel 2006 ja 2007 ETU-s vastusevariantide hulgas ka „Ei tea“, mistõttu nendel aastatel on teadmata väärtuste hulk märksa suurem kui järgmistel aastatel (2006. a uuringus teadmata 454 isikul, 2007. a 545 isikul, kuid 2008. a 94 isikul).
- Perekonnaseis oli 2006 ja 2007. aasta ETU-s vastamata neil, kes olid uuringu ajal alla 16-aastased, sest seda neilt ankeedis ei küsitud. Ühendandmebaasis on need isikud kodeeritud vastusevariandi „Vallaline“ alla.
- Aastatel 2006 ja 2007 ei olnud ESU-s eraldi küsimust juhtimiskohustuste kohta, nagu on olnud järgmistel aastatel. Kuna nendel aastatel on aga isikuankeedis küsimus isiku alluvuses töötavate isikute arvu kohta, on kasutatud nende aastate andmebaasides just seda tunnust. Kui respondendil oli vähemalt üks tema otseses alluvuses töötav isik, omistati talle juhtimiskohustuste olemasolu.

Aastatel 2006–2008 olid ETU andmebaasid teisel kujul kui hilisemad. Kui tavaliselt on ühe aasta uuringuandmed kolmes failis (leibkonna, leibkonnaliikme ja isiku tasandil), siis nimetatud aastate andmed on koos ühes failis, kus vastutava leibkonnaliikme juures on ka teiste leibkonnaliikmete andmed. Selleks et siduda leibkonnaliige isikufailiga, tuleb kasutada identifitseerivat tunnust ehk isiku unikaalset koodi. Ühtlasi tuleb silmas pidada, et ETU-s küsitletakse sama leibkonda kahes järjestikuses kvartalis ehk üks leibkond on pooleteise aasta jooksul uuringus neli korda.

Alates 2009. aastast on ETU ja ESU andmebaasid sarnasemad, kuna isiku-uuringutes ühtlustati tausttunnuseid ja eri uuringute tausttunnused harmoneeriti. Tausttunnuste ühtlustamist on täpsemalt käsitletud „Tausttunnuste käsiraamatus“^a.

2010. aastal läks Statistikaamet üle uuele ametite klassifikaatorile ISCO-08 (International Standard Classification of Occupations). Kuna uus ja vana klassifikaator üks-üheselt ei kattu, on aastatel 2006–2009 ametialad kodeeritud ümber uue, praegu kasutusel oleva klassifikaatori järgi. Nii uus kui ka eelmine ametialade klassifikaatori versioon on kättesaadavad Statistikaameti veebilehel^b, kus on ühtlasi klassifikaatori eri versioonide üleminekutabelid^c. Kuna ESU andmed ei olnud tagasiulatavalt ümber kodeeritud, tehti seda aastate 2006–2009 kohta tagantjärele.

Palgatunnuse kvaliteet isiku-uuringutes

ETU ja ESU palgatunnuse kvaliteeti on põhjalikult analüüsitud kogumikus „Palgaerinevuste statistika parem kättesaadavus: kasutatavad andmeallikad“^d.

Maksu- ja Tolliameti andmete lisamine ühendandmebaasi

Maksu- ja Tolliameti andmebaasi struktuur

MTA andmebaas tugineb TSD-vormi lisaga 1 kogutud andmetele. TSD-vorm on tulu- ja sotsiaalmaksu, kohustusliku kogumispensioni makse ja töötuskindlustusmakse deklaratsioon, mille esitab MTA-le väljamakse tegija väljamaksele järgneva kalendrikuu 10. kuupäevaks. Deklaratsiooni vorm lisadega ja nende täitmise kord on kehtestatud rahandusministri määrusega.

MTA andmebaasis on järgmised tunnused:

- väljamakse saaja isikukood;
- väljamakse tegija registrikood;
- väljamakse tegemise aasta;
- väljamakse tegemise kuu;
- väljamakse suurus liigi järgi.

^a <http://www.stat.ee/dokumendid/31703>

^b <http://metaweb.stat.ee/?siteLanguage=ee>

^c http://metaweb.stat.ee/classificator_correspondence_list.htm?siteLanguage=ee

^d <http://www.stat.ee/dokumendid/78174>

Väljamakse saaja isikukoodist leitakse Statistikaametis anonümiseerimise teel väljamakse saaja sugu ja sünniaeg ning luuakse anonüümne isikukood, mille kaudu on võimalik registripõhiseid andmeid teiste andmeallikatega ühitada.

MTA andmebaasis ei vasta igale isikule alati üks kirje. Esialgses MTA andmebaasis tähendab üks kirje ühelt ettevõttelt ühes kuus ühele isikule tehtud väljamakseid. MTA andmebaasis esineb järgmisi kirjeid:

- kui isik on saanud ühes kuus väljamakseid mitmelt tööandjalt, on iga ettevõtte väljamaksed eri kirjetes;
- kui sama tööandja on ühele isikule teinud ühes kuus eri liiki väljamakseid, kajastub see andmebaasis ühe kirjena;
- kui isik on saanud töötasu samalt tööandjalt eri kuudes, on andmebaasis iga kuu kohta üks kirje;
- kui isik on saanud eri liiki väljamakseid samalt tööandjalt eri kuudes, on iga kuu kohta üks kirje.

MTA andmetes kajastatud väljamakseliikide alusel on koostatud statistiline näitaja „tõine tulu“, mille juures võetakse arvesse järgmisi väljamakse liike:

- väljamakse liik nr 1: töötasu;
- väljamakse liik nr 11: juhatuse liikme tasu;
- väljamakse liik nr 15: võlaõigusliku lepingu alusel saadud töötasu (tavapäraselt töövõtulepinguga töötamise puhul).

Tõise tulu puhul lähevad arvesse ka negatiivsed väljamaksed.

Maksu- ja Tolliameti andmete lisamine tööjõu-uuringu ja sotsiaaluuringu ühendandmebaasi

Selleks et MTA andmed ETU ja ESU ühendandmebaasi omadele liita, tuleb MTA andmebaas kõigepealt isiku tasandile viia, s.t iga isiku kohta jääb ainult üks kirje.

Kuna MTA andmebaasi põhjal koostatakse ka mitmesugused tunnused kolme kuu keskmise tõise tulu kohta (uuringukuu ning sellele eelnev ja järgnev kuu), on vajalik iga aasta puhul arvestada ka eelmise aasta viimase kuu ja järgmise aasta esimese kuu väljamakseid. Näiteks 2012. aasta andmebaasis on kõik 2012. aastal tehtud väljamaksed, 2011. aasta detsembris tehtud väljamaksed ja 2013. aasta jaanuaris tehtud väljamaksed. Kuna MTA andmebaasis on kõik isikud, kes on MTA andmetel väljamakseid saanud, jäetakse alles ainult need, kes osalesid vastaval aastal uuringus (kas ETU-s või ESU-s). Seejärel leitakse iga kirje puhul kolm tõise tulu liiki ja arvutatakse muu tulu kokku. Selle tulemusena saadakse vastava aasta MTA andmebaas.

Et MTA andmebaas isiku tasandile viia, tuleb andmebaasi kirjed koondada (kõigepealt koondatakse isiku, kuu ja aasta tasandil). See tähendab, et iga isiku puhul jääb iga kuu kohta üks kirje. Kui isik on saanud ühes kuus väljamakseid mitmelt tööandjalt, siis ühe kuu tõise tulu liigid summeeritakse. Seejärel restruktureeritakse andmebaas nii, et iga isiku kohta on üks rida ning iga kuu väljamaksed on esitatud eraldi tunnusena.

MTA andmebaasis olevad palgaandmed on võimalik lisada kõikidele, kes on osalenud ETU-s või ESU-s ja kelle isikukood on teada. Tabelis 8 on teadmata isikukoodiga isikute arv aastatel 2006–2013 ehk need, kellele ei õnnestunud MTA andmebaasist nende palgaandmeid juurde linkida.

Tabel 8. Teadmata isikukoodiga isikud ühendandmebaasis

	2006	2007	2008	2009	2010	2011	2012	2013
Isikukood teadmata	4 448	449	742	314	329	286	193	150

Tabelis 9 on välja toodud, kui palju on ETU ja ESU andmetel palgatöötajaid ning kui paljudel oli MTA andmete põhjal teine tulu suurem kui null.

Tabel 9. Palgatöötajad ja teine tulu uuringu järgi, 2006–2013

Aasta	ETU	ESU	Kokku
2006			
Palgatöötajad	9 005	6 419	15 424
Töine tulu olemas	6 245	6 109	12 354
2007			
Palgatöötajad	10 513	6 033	16 546
Töine tulu olemas	9 792	5 679	15 471
2008			
Palgatöötajad	10 165	5 453	15 618
Töine tulu olemas	9 243	5 142	14 385
2009			
Palgatöötajad	8 033	5 104	13 137
Töine tulu olemas	7 482	4 819	12 301
2010			
Palgatöötajad	7 828	4 708	12 536
Töine tulu olemas	7 294	4 443	11 737
2011			
Palgatöötajad	8 970	4 887	13 857
Töine tulu olemas	8 365	4 577	12 942
2012			
Palgatöötajad	10 542	5 363	15 905
Töine tulu olemas	9 893	5 054	14 947
2013			
Palgatöötajad	10 605	5 753	16 358
Töine tulu olemas	9 975	5 418	15 393

Maksu- ja Tolliameti andmebaasi põhjal koostatud tunnused**Töine tulu**

Pärast andmebaasi viimist isiku tasandile leitakse vajalikud tunnused, mis on seotud sissetuleku liikide, teise tulu ja põhitööandja väljamakstud tasuga. Ühendandmebaasis olevad tunnused on esitatud kogumiku lõpus (vt lisa 1).

Esmalt leitakse tunnused teise tulu kohta kokku ja eraldi selle eri liikide kohta (S_1 ehk töötasu, S_11 ehk juhatause liikme tasu, S_15 ehk võlaõigusliku lepingu alusel saadud töötasu ja MT ehk muu tulu). Iga aasta andmebaasis on arvatud teine tulu nende kolme TSD-liigi kohta (eelmise aasta detsember, vaadeldav aasta ja järgmise aasta jaanuar). Töine tulu iga kuu kohta saadakse nende kolme väljamakseliigi summeerimisel.

Tunnuste S_1, S_11 ja S_15 summa on teine tulu kokku, mis on leitud samuti eelmise aasta detsembri (TT_12_EA), vastava aasta iga kuu (TT_01 ... TT_12) ja järgmise aasta jaanuari (TT_01_JA) kohta. Samuti on arvatud kõikide sissetulekuliikide ja teise kogutulu aastasumma, mis on kuiste töötulude summa.

Et teada saada, mitu kuud on isik töist tulu saanud, liidetakse kokku nende kuude arv, millal ta on seda on saanud (s.t isik on saanud vähemalt ühte kolmest väljamakse liigist). Kui isik on saanud mõnes kuus negatiivseid väljamakseid, võetakse ka see kuude arvutamisel arvesse. Samuti leitakse kolme väljamakseliigi ja muude tulude puhul nende kuude arv, millal vastavat liiki väljamakset saadi.

Põhitööandja makstud töine tulu

Kõik töise tulu kohta arvatud tunnused leitakse ka põhitööandjale, kes on suurima väljamakse teinud tööandja.

Põhitööandja on määratud järgmiselt.

- Kõigepealt määratakse ühe kuu põhitööandja: kui isik on saanud ühes kuus väljamakseid mitmelt tööandjalt, on põhitööandja see, kes on teinud töise tulu väljamakse kõige suuremas summas ehk kelle tunnuste S_1 + S_11 + S_15 summa on suurim.
- Aasta põhitööandja määratakse kuu põhitööandjate põhjal ja arvesse võetakse ainult kalendriaasta ehk 12 kuud. Põhitööandja on see, kes on teinud väljamakseid pikema aja jooksul (suurem kuude arv). Kui mitu tööandjat on teinud väljamakseid sama pika aja jooksul, on põhitööandja see, kes on teinud töise tulu väljamakse suuremas summas (tunnuste S_1 + S_11 + S_15 summa on suurim).

Pärast põhitööandja määramist leitakse põhitööandja tehtud väljamaksed ja nende tunnuste ees on kasutatud lühendit PT (põhitööandja). Iga aasta andmebaasis on eraldi leitud sellele eelnenud aasta detsembri, vastava aasta iga kuu ja sellele järgnenud aasta jaanuari kohta põhitööandja väljamakstud töise tulu liigid (PT_S1, PT_S11, PT_S15) ja muu tulu kokku (PT_MT). Samuti on arvatud põhitööandja väljamakstud töine tulu vastava aasta iga kuu (PT_01 ... PT_12), sellele eelnenud aasta detsembri (PT_12_EA) ja sellele järgnenud aasta jaanuari (PT_01_JA) kohta.

Peale igakuiste väljamaksete on leitud ka põhitööandja makstud töine tulu aastas kokku (PT_aasta) ja nende kuude arv, millal seda saadi (PT_kuude_arv). Samuti on iga väljamakseliigi kohta eraldi arvatud väljamakstud summa aastas ja nende kuude arv, millal seda välja maksti. MTA andmete põhjal koostatud andmebaasi tunnuste loetelu on esitatud lisa 1.

Töise tulu põhjal arvatud lisatunnused

Peale eespool kirjeldatud töise tulu väljamaksete ja põhitööandja tehtud väljamaksete tunnuste on andmebaasis ka mitmesugused lisatunnused (vt lisa 1). Eespool kirjeldatud tunnuste põhjal on arvatud keskmine töine tulu kuus kalendriaastas ja keskmine töine tulu tunnis.

Aastakeskmise töise tulu leidmiseks jagatakse kogu aasta töine tulu nende kuude arvuga, millal töist tulu saadi. Aastakeskmise töise tulu puhul on arvesse võetud ainult need isikud, kelle aastakeskmise töine tulu oli suurem kui null.

Valem selle leidmiseks on järgmine:

$$\text{Aastakeskmise töine tulu kuus} = \frac{\text{Töine tulu aastas kokku}}{\text{Nende kuude arv, kui töist tulu aastas saadi}}$$

Andmebaasis tunnuste nimetustega on valem järgmine:

$$TT_{ak} = \frac{TT_{aasta}}{TT_{kuude_arv}}$$

Kuna ETU-s ja ESU-s küsitletakse inimesi eri kuudes, ei pruugi aastakeskmise töine tulu kuus minna kokku selle töökoha palgaga, kus isik uuringu ajal parajasti töötas. Näiteks võis ta uuringu tegemise ajal töötada madalama palgaga töökohal, kuid hiljem ülejäänud aasta kõrgema palgaga töökohal, mistõttu võib tema aastakeskmise töine tulu uuringus öeldud sissetulekust suurel määral erineda. Selle väljaselgitamiseks on ühendandmebaasis järgmised töise tulu tunnused:

- töine tulu uuringukuus (TT_uk) on uuringu tegemise kuu järgi vastava kuu töine tulu. Näiteks kui isikut küsitleti jaanuaris, arvestatakse talle jaanuaris tehtud väljamaksete summat;

- uuringukuule järgneva kuu tõine tulu (TT_ujk) on uuringu tegemisele järgnenud kuu tõise tulu summa. Näiteks kui isikut küsitleti jaanuaris, arvestatakse talle veebruaris tehtud väljamaksete summat;
- kolme kuu keskmise tõise tulu arvutamine (uuringukuu ning sellele eelnev ja järgnev kuu) (TT_3k) koosneb kahest etapist. Esmalt leitakse iga uuringukuu puhul kolme kuu tõine tulu kokku (uuringukuule eelnev kuu, uuringukuu ja uuringukuule järgnev kuu). Näiteks jaanuari puhul liidetakse kokku eelmise aasta detsembri ning vastava aasta jaanuari ja veebruari tõise tulu väljamaksed. Seejärel leitakse, mitmel kuul vaatlusalusest kolmest kuust isik tõist tulu sai. Kui ta on saanud vastavas kuus negatiivseid väljamakseid, arvestatakse ka need tõise tulu hulka. Kolme kuu keskmise tõise tulu saamiseks jagatakse kolme kuu tõine tulu nende kuude arvuga, millal tõist tulu saadi.

Analüüsi tulemusena otsustati tõise tuluna kasutada isiku aastakeskmist tõist tulu kuus (TT_ak). Täpsem analüüs on kättesaadav kasutatavate andmeallikate ühilduvusanalüüsis^a (lk 55–66).

Peale aasta- ja kuukeskmise tõise tulu on võimalik arvutada ka tõine tulu tunnis. Selleks saab kasutada nii inimese uuringus öeldud tundide arvu (tavaline nädalas töötatud tundide arv) kui ka ettenähtud töötundide arvu. Tunnipalga leidmiseks on vaja teada ka tööpäevade arvu aastas. Aasta tööpäevade arvuks on võetud 253, mis on aastate 2006–2014 keskmine. Keskmist arvu kasutatakse sellepärast, et palk ei muutu ka juhul, kui aastas on tööpäevi sellest rohkem või vähem.

Tabel 10. Tööpäevi aastas, 2006–2014

	2006	2007	2008	2009	2010	2011	2012	2013	2014	Keskmine
Tööpäevade arv	254	254	253	252	254	255	254	253	251	253,333

Tunnipalka arvutatakse järgmise valemiga:

$$\text{Tunnipalk} = \frac{\text{Aastakeskmise tõine tulu kuus} * 12}{(\text{Tavaline töötatud tundide arv nädalas} / 5) * 253}$$

Andmebaasis olevate tunnuste nimetustega on valem järgmine:

$$\text{TT_tund} = \frac{\text{TT_ak} * 12}{(\text{T_tund} / 5) * 253}$$

Tunnipalga arvutamiseks ettenähtud tundide järgi asendatakse valemis tavaline nädalas töötatud tundide arv ettenähtud töötundide arvuga. Ettenähtud töötunnid arvutatakse järgmiselt:

- kui isik töötab osalise tööajaga ja tavaline nädalas töötatud tundide arv on alla 35, siis ettenähtud töötunnid = tavaline nädalas töötatud tundide arv;
- kui isik töötab osalise tööajaga ja tavaline nädalas töötatud tundide arv on vähemalt 35, siis ettenähtud töötunnid = 20 tundi;
- kui isik töötab täistööajaga ja on haridustöötaja, siis ettenähtud töötunnid = 35 tundi;
- kui isik töötab täistööajaga, kuid ei ole haridustöötaja, siis ettenähtud töötunnid = 40 tundi.

Laiendustegurite (kaalude) arvutamine

Et ETU ja ESU ühendandmebaasi tulemusi saaks laiendada üldkogumile, tuleb arvutada uued laiendustegurid ehk kaalud. Ühtlasi elimineerib see ETU-s korduvate isikute topeltarvestuse probleemi, sest võimaldab korduvad isikud võtta arvesse kui ühel korral uuringus osalenud isikud.

^a <http://www.stat.ee/dokumendid/78174>

Leibkonnakaalud

Ühise andmebaasi tarvis on kaalud arvutatud uuringuspetsiifiliste kaalude põhjal. ETU-s on kasutatud kaalu nimetusega WGT_Y, ESU-s kaalu nimetusega DB090. Andmestike kokkupanemisel on ESU kaal korrutatud kordajaga λ ja ETU kaal kordajaga $1-\lambda$. Kordaja λ väärtus oleneb uuringute efektiivsete valimimahtude suhtest:

$$\alpha = \frac{n_{ESU}^{eff}}{n_{ESU}^{eff} + n_{ETU}^{eff}},$$
$$n_{ESU}^{eff} = \frac{n}{1 + \sqrt{\text{var}(DB090)/\text{mean}(DB090)}},$$

kus n on ESU leibkondade arv;

$$n_{ETU}^{eff} = \frac{n}{1 + \sqrt{\text{var}(WGT_Y)/\text{mean}(WGT_Y)}},$$

kus n on ETU leibkondade arv.

Suurema kaalu saab uuring, millel on suurem valimimaht ja seega väiksem kaalude varieeruvus. Näiteks 2011. aastal $\lambda = 0,358$ ja 2012. aastal $\lambda = 0,346$. Tänu sellele on eri uuringutest tulevate leibkondade kaalud ühises andmestikus sarnased.

Lõplikud kaalud on kalibreeritud kümne aasta soo- ja vanuserühmade ning maakonna ja asustusüksuse liigi järgi. Kasutatud on korrigeeritud rahvaarvu, millest on lahutatud institutsioonides elavate isikute arv.

Isikukaalud

Kui analüüsis kasutatakse ainult isikuankeedi täitnud isikuid, kasutatakse isikukaalusid, mille arvutamisel on arvestatud leibkonnasisest vastamata jätmist. Need kaalud on arvutatud leibkonnakaalude põhjal nii, et ühe aasta soo- ja vanuserühmades on kaalude summad vastanute ja kõikide vähemalt 15-aastaste isikute puhul võrdsed. Seega isikukaalud leibkonna sees enam võrdsed ei ole.

Et ühendandmebaasis oleval tunnusel on saadud isikuankeedist, on ühendandmebaasile lisatud ainult isikukaal.

Loodud ühendandmebaasis on kokku kaks kaalutunnust: uuringukaal ja kaal. Andmebaasi lisatud uuringukaal võimaldab uuringute andmeid vaadelda uuringute kaupa eraldi. Kaalutunnus on aga ühendandmebaasi peale ühtne, mis võimaldab teha analüüse kogu ühendandmebaasi põhjal.

Puuduvate väärtuste imputeerimine

Mõned tunnused, nagu palk, töötatud tundide arv ja tööaja vorm, on palgalõhe arvutamisel väga olulised. Selleks et vähendada nendes küsimustes teadmata väärtuste osatähtsust, leitakse puuduvatele väärtustele nende võimalik prognoos ehk puuduolevad väärtused imputeeritakse.

Palga imputeerimine Eesti tööjõu-uuringus

ETU-s küsitakse viimase kuu täpset töötasu kõikidelt palgatöötajatelt. Et ei tekiks segadust brutoja/või netotasu eristamisega, küsitakse mõlemat. Neile, kes ei oska või ei taha öelda oma täpset töötasu, palutakse öelda tema viimase kuu netokuutöötasu vahemik. Nendele vastajatele, kes oma palka uuringus ei avalda, leitakse see teiste tunnuste alusel ehk nendele isikutele netokuupalk imputeeritakse.

Esmalt imputeeritakse netokuupalka vahemik kõigile neile palgatöötajatele, kes seda öelda ei tahtnud või ei osanud. Vahemik imputeeritakse juhusliku logistilise regressiooniga tarkvaras IVEWARE. Tunnustena on mudelis kasutatud isiku sugu, vanuserühma (viie aasta vanuserühmad), elukoha maakonda ja elukoha asustusüksuse liiki.

Seejärel imputeeritakse palgavahemiku järgi kõigile palgatöötajatele täpne netokuupalk *hot-deck*-meetodiga, kus juhuslikult valitakse isik samast palgavahemikust ja kopeeritakse tema täpne palgaväärtus.

Kuna imputeerimisel kasutatakse netopalka, on ühendandmebaasis kõikide palgatöötajate kohta olemas netopalk. Ühendandmebaasis kasutatav brutopalk leitakse järgmise valemiga:

kui netopalk ≤ 144 , siis brutopalk = netopalk

kui netopalk > 144 , siis brutopalk = $(\text{netopalk} - 0,21 \cdot 144) / (1 - 0,21) = (\text{netopalk} - 30,24) / 0,79$

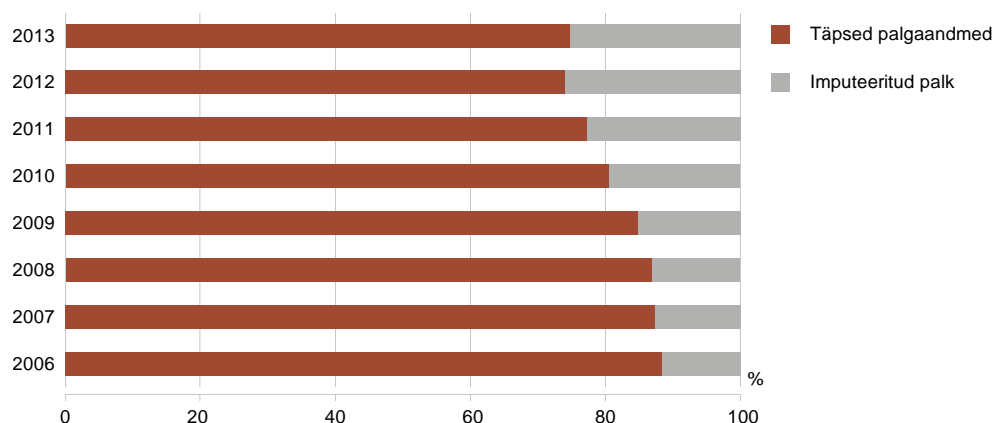
Täpsem ülevaade netopalka brutopalgaks ümberarvutamise kohta eri aastatel on tabelis 11. Palgateisendustel on arvestatud üksnes tulumaksu. Aastate 2006–2010 summad on kroonides, aastate 2011–2013 omad eurodes.

Tabel 11. Netopalka ümberarvutamine brutopalgaks, 2006–2013

Aasta	Maksuvaba, krooni/eurot	Maksumäär, %	Maksumäär * 1 – maksumäär	Valem
2006	2 000	24	480,0	0,76 $B = (N - 480) / 0,76$
2007	2 000	22	440,0	0,78 $B = (N - 440) / 0,78$
2008	2 250	21	472,5	0,79 $B = (N - 472,5) / 0,79$
2009	2 250	21	472,5	0,79 $B = (N - 472,5) / 0,79$
2010	2 250	21	472,5	0,79 $B = (N - 472,5) / 0,79$
2011	144	21	30,2	0,79 $B = (N - 30,24) / 0,79$
2012	144	21	30,2	0,79 $B = (N - 30,24) / 0,79$
2013	144	21	30,2	0,79 $B = (N - 30,24) / 0,79$

Joonisel 2 on välja toodud, kui suur osa uuringus osalenud palgatöötajatest ütles oma täpsed palgaandmed ja kui suurel osal on palk imputeeritud.

Joonis 2. Tööjõu-uuringu palgaandmed, 2006–2013



Palga imputeerimine Eesti sotsiaaluuringus

ESU-s küsitakse viimase kuu töötasu samuti kõikidelt palgatöötajatelt. Põhitöökohal saadava kuupalka puhul küsitakse vaid netosummat, brutoväärtus arvutatakse hiljem töötluse käigus. Palgatunnused imputeeritakse nende isikute puhul, kes keelduvad oma palka ütlemast või ei tea seda.

Lõpuks jäi keskmise netokuupalka kohta vastamata nendel, kes ei soovinud või ei osanud palgaküsimusele vastata. Neile leitakse netopalk statistiliste meetoditega teiste tunnuste kaudu ehk nende netopalk imputeeritakse.

Palgatunnus imputeeritakse juhusliku lineaarse regressiooni abil tarkvaraga IveWare ja kasutatavas imputeerimismudelil on järgmised tunnused:

- eelmise aasta keskmine kuupalk ja selle kvartiil;
- isiku vanus, vanuse ruut ja sugu;
- näitaja selle kohta, kas isik töötas eelmisel aastal vähemalt 11 kuud;
- näitaja selle kohta, kas isikul on juhtimiskohustused;
- näitaja selle kohta, kas isik vahetas viimase 12 kuu jooksul töökohta;
- näitaja selle kohta, kas üksuses, kus isik töötab, on üle 10 töötaja;
- näitaja selle kohta, kas isiku amet kuulub teenindus- ja müügitöötajate rühma.

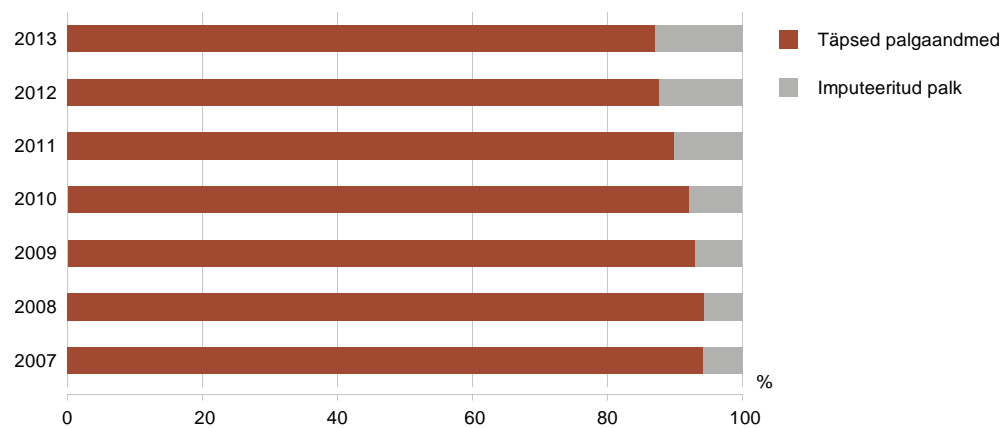
Kuna ühendandmebaasi otsustati lisada brutopalgatunnus, teisendatakse imputeeritud netopalk seejärel brutopalgaks järgmiselt:

kui netopalk ≤ 144 , siis brutopalk = netopalk

kui netopalk > 144 , siis brutopalk = $(\text{netopalk} - 0,21 \cdot 144) / (1 - 0,21) = (\text{netopalk} - 30,24) / 0,79$

Täpsem ülevaade netopalka brutopalgaks ümberarvutamise kohta eri aastatel on tabelis 11.

Joonis 3. Sotsiaaluuringu palgaandmed, 2007–2013



Töötatud tundide arvu imputeerimine

Väljundnäitajate analüüsimisel selgus, et töötatud tundide arvu puhul esines andmebaasis ebaloogilisi väärtusi. Seetõttu otsustati ebaloogilistena tunduvate tundide asemele imputeerida uued väärtused. Selleks arvutati Eesti keskmine brutotunnipalk tegevusalade järgi ja võrreldi seda uuringus öeldud tundide arvu põhjal arvutatud tunnipalgaga.

Kõigepealt leiti Statistikaameti andmebaasi andmete põhjal Eesti keskmine brutotunnipalk põhi-tegevusala (EMTAK 2008) järgi aastate 2008–2013 kohta^a eraldi nii täis- kui ka osalise töötajaga töötajatel. Seejärel on leitud täis- ja osalise töötajaga töötajate keskmisest brutotunnipalgast eri osatähtsused: 10%, 20% või 30%. Nende osatähtsuste järgi arvutati tegevusalade keskmise brutotunnipalga vahemikud, mille piires oli lubatud uuringu brutotunnipalgal varieeruda. Tulemusi analüüsid otsustati, et tunnipalk võib keskmisest brutotunnipalgast erineda pluss-miinus 30%. Neid piire kasutati selleks, et teha kindlaks, kas töötatud tundide arv on vaja imputeerida või mitte.

Et 2006. ja 2007. aastal ei olnud klassifikaator veel EMTAK 2008, vaid selle vanem versioon^b, tuli nende kahe aasta puhul tunnipalk eraldi arvutada. Aastate 2006 ja 2007 kohta oli teada tegevusalade keskmine brutotunnipalk, seega leiti kõigepealt koefitsient, mis näitab ühe aasta ja sellele järgnenud aasta tegevusalade keskmise brutotunnipalga vahe suhet järgnenud aasta brutotunnipalka. Seejärel

^a Statistikaameti andmebaas, tabel PA5212: Keskmine brutotunnipalk põhitegevusala (EMTAK 2008) järgi

^b Statistikaameti andmebaas, tabel PA5312: Keskmine brutotunnipalk põhitegevusala (EMTAK 2003) järgi (2000–2008)

lahutati järgmise aasta vastava tegevusala brutotunnipalgast koefitsiendiga läbi korrutatud brutotunnipalk. Näiteks 2008. aastal oli täistööajaga töötajate keskmine brutotunnipalk 5,02 eurot ja 2007. aastal 4,36 eurot. Koefitsiendi väärtus on $(5,02 - 4,36) / 5,02 = 0,13$. Põllumajanduse, metsamajanduse ja kalapüügi tegevusalal oli täistööajaga töötajate keskmine brutotunnipalk 2008. aastal 4,06 eurot. Sama tegevusala 2007. aasta keskmine brutotunnipalk oli seega $4,06 - (4,06 * 0,13) = 3,53$ eurot. 2006. aasta kohta arvatati kõigepealt samuti koefitsient ja seejärel eri tegevusalade keskmine brutotunnipalk võrreldes 2007. aastaga. Lisas 2 on välja toodud täis- ja osalise tööajaga töötajate brutotunnipalga 30% vahemikud tegevusala kaupa.

Isikud, kellele töötatud tundide arv imputeeritakse, on leitud kahel viisil. Üks osa küsitletavatest vastas, et neil on raske öelda oma tavalist nädalas töötatud tundide arvu, sest see erineb nädalati või kuuti väga palju. Neile, kes nii vastasid, töötatud tundide arv imputeeriti. Teine osa leiti eespool arvatud brutotunnipalga 30% vahemikke kasutades. Selle kohaselt imputeeriti töötatud tunnid juhul, kui isiku tunnipalk (MTA andmete põhjal) ei langenud keskmise brutotunnipalga 30% vahemikku, s.t brutotunnipalk oli sellest suurem või väiksem. Töötatud tundide arvu arvestati juhul, kui see oli kõige rohkem 10 või vähemalt 60. Seega töötatud tundide arv imputeeriti juhul, kui olid täidetud järgmised tingimused:

- isik oli töötav ($I_TOOTAV = 1$);
- isiku töötatud tundide arv oli 0 ja 10 vahel ($T_TUND > 0 \ \& \ T_TUND \leq 10$) või vähemalt 60 ($T_TUND \geq 60$);
- isiku brutotunnipalk ei langenud Eesti keskmise brutotunnipalga 30% vahemikku.

Selle tõttu leidus isikuid, kellele oli vaja töötatud tundide arv uuesti imputeerida. Arvestades neid ja ka neid, kes keeldusid töötatud tundide arvu ütlema, on tabelis 12 esitatud nende isikute koguarv, kellele tuli töötatud tundide arv uuesti imputeerida.

Tabel 12. Töötatud tundide arvu imputeerimine, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Imputeeritud tundide arvuga isikud	283	268	260	245	298	383	517	494

Kõigepealt imputeeritakse töötundide vahemikud juhusliku logistilise regressiooniga tarkvaraga IVEWARE ja seejärel täpsed väärtused iga vahemiku sees *hot-deck*-meetodiga. Vahemike imputeerimisel on abitunnused ameti- ja tegevusalarühmade näitajad, kõrvaltöö olemasolu, vanuserühm, teine tulu aastas kokku, tööalane staatus, palgatöötaja/mittepalgatöötaja, juhtimiskohustuste olemasolu.

Ühendandmebaasi tunnustes on välja toodud esialgne (uuringupõhine) tavaline nädalas töötatud tundide arv (T_TUND), imputeeritud väärtustega töötatud tundide arv (T_TUND_imp) ja tunnus, mis näitab, kas töötatud tundide arv on imputeeritud või mitte (t_tund_i).

Töötaja vormi imputeerimine

Nii nagu töötatud tundide arvu puhul leidus ka töötaja vormi (täis- või osalise tööajaga töötamine) puhul andmebaasis ebaloogilisi tulemusi.

Selleks et teada saada, kui paljudele isikutele on vaja töötaja vorm imputeerida, summeeritakse need, kelle puhul on juba ETU või ESU andmebaasis töötaja vorm teadmata, ja need, kelle töötaja vorm tundub ebaloogiline. Seega imputeeriti töötaja vorm juhul, kui olid täidetud järgmised tingimused:

- isik töötas ($I_TOOTAV = 1$);
- isik töötas täistööajaga ja töötatud tundide arv oli vahemikus 0 kuni 10 ($T_AEG = 1$ & ($T_TUND > 0$ & $T_TUND \leq 10$)) või isik töötas osalise tööajaga ja töötatud tundide arv oli vähemalt 60 ($T_AEG = 2$ & $T_TUND \geq 60$).

Tabelis 13 on välja toodud, kui paljudele isikutele aastatel 2006–2013 tööaja vorm imputeeriti.

Tabel 13. Tööaja vormi imputeerimine, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Imputeeritud tööaja vormiga isikud	0	2	6	7	7	14	15	20

Tööaja vorm imputeeritakse juhusliku logistilise regressiooni teel, kasutades tarkvara IVEWARE. Tunnuse T_AEG imputeerimisel on abitunnused ameti- ja tegevusalarühmade näitajad, kõrvaltöö olemasolu, vanuserühm, töine tulu aastas kokku.

Ühendandmebaasi tunnustes on välja toodud esialgne (uuringu põhjal) tööaja vorm (T_AEG), imputeeritud väärtustega tööaja vorm (T_AEG_imp) ja tunnus, mis näitab, kas tööaja vorm on imputeeritud või mitte (t_aeg_i).

Ühendandmebaasi loomise kokkuvõte

Projekti „Palgaerinevuste statistika parem kättesaadavus“ raames loodi ühendandmebaasid aastate 2006–2013 kohta – kokku 8 andmebaasi. Liideti kahe isiku-uuringu – Eesti tööjõu-uuringu ja Eesti sotsiaaluuringu – andmebaasid. Neile lisati veel andmeid Maksu- ja Tolliameti andmebaasist. Et andmeid oleks võimalik laiendada üldkogumile, leiti ühendandmebaasidele uued laiendustegurid ehk kaalud.

Lõplikus ühendandmebaasis on kokku 246 tunnust: 61 tunnust, mis tuginevad ETU-le ja ESU-le, 160 tunnust, mis tuginevad MTA andmetele, ning 25 lisatunnust, mis on leitud nii isiku-uuringute kui ka MTA andmete põhjal (vt lisa 1).

PALGALÕHE JA SEDA MÕJUTAVAD NÄITAJAD

Projekti „Palgaerinevuste statistika parem kättesaadavus“ raames avaldatakse sooline palgalõhe eri tausttunnuste järgi ja mitmesugused lisanäitajad. Nende näitajate aluseks on võetud poliitikauuringute keskuse PRAXIS ja Eesti rakendusauuringute keskuse CentAR antud soovituslikud näitajad. Et mõista soolise palgalõhe dünaamikat, on samal ajal vaja jälgida, kuidas arenevad ühiskond ja tööturg nendes aspektides, mis on soolise palgalõhe tekke ja püsimise taga^a. Lõplik näitajate loetelu ja nende arvutamise meetodika lepiti ekspertidega kokku 2015. aasta märtsis Statistikaametis toimunud töötoas. Töötoa materjalid on kättesaadavad Statistikaameti veebilehel^b.

Järgnevalt on esitatud näitajad koos nende lühikirjeldusega.

Keskmine sooline palgalõhe

Näitab meeste ja naiste tunnipalga erinevuse suhet meeste tunnipalka. Arvutusvalem:

$$\frac{\text{meeste brutotunnipalk} - \text{naiste brutotunnipalk}}{\text{meeste brutotunnipalk}} \times 100\%$$

Andmeallikas: ühendandmebaas

Ajavahemik: 2006–2013

Ühik: %

Soolist palgalõhet arvutatakse järgmiste tausttunnuste põhjal:

- vanus
- rahvus
- elukoht
- haridustase
- kõrgeima omandatud hariduse eriala
- ametiala
- tegevusala
- tööaja vorm
- juhtimiskohustuste olemasolu
- tööandja liik
- ettevõtte/asutuse suurus
- tegelik perekonnaseis
- eelkooliealiste laste olemasolu

Osalise tööajaga töötajate osatähtsus nais- ja meessoost palgatöötajate hulgas

Näitab, kui suur osa mehi ja naisi töötab osalise ning kui suur osa täistööajaga.

Andmeallikas: tööjõu-uuring

Ajavahemik: 2006–2013

Ühik: %

^a https://www.sm.ee/sites/default/files/content-editors/Ministeerium_kontaktid/Valjaanded/toimetised_20112.pdf

^b <http://www.stat.ee/183300>

Sooline segregatsioon tööturul

Näitab, kui suur osa naistest peaks töökohta vahetama, et naiste ja meeste osatähtsus oleks kõikidel ameti- ja tegevusaladel võrdne. Arvutamiseks kasutatakse Duncani ja Duncani indeksi. Segregatsiooniindeksi saab arvutada nii ameti- kui ka tegevusala põhjal.

Arvutusvalem:

$$S_t = \frac{1}{2} \sum_i \left| \frac{q_{it} T_{it}}{\sum_i q_{it} T_{it}} - \frac{p_{it} T_{it}}{\sum_i p_{it} T_{it}} \right|$$

Kus $p_{it} = F_{it} / T_{it}$ ja $q_{it} = (1 - p_{it}) = M_{it} / T_{it}$

F_{it} = naiste arv tegevusalal i ja aastal t

M_{it} = meeste arv tegevusalal i ja aastal t

$T_{it} = F_{it} + M_{it}$ = töötajate koguarv tegevusalal i ja aastal t

Lühendatud kujul:

$$S_t = \frac{1}{2} \sum_i \left| \frac{M_i}{M} - \frac{F_i}{F} \right|$$

Andmeallikas: ühendandmebaas

Ajavahemik: 2006–2013

Ühik: %

Meeste ja naiste osatähtsus juhtivtöötajate hulgas

Näitab, kui suur osa juhtivtöötajatest on mehed ja kui suur osa naised. Juhtivtöötajateks arvestatakse need isikud, kes on ISCO järgi juhid ja kellel on juhtimiskohustus.

Andmeallikas: ühendandmebaas

Ajavahemik: 2006–2013

Ühik: %

Meeste ja naiste osatähtsus ettevõtjate hulgas

Näitab, kui suur osa ettevõtjatest on mehed ja kui suur osa naised. Ettevõtjad on need hõivatud, kes määratlevad end üksikettevõtja, palgatöötaja(te)ga ettevõtja või palgatöötajana talupidajana.

Andmeallikas: ühendandmebaas

Ajavahemik: 2006–2013

Ühik: %

Lapsehooldusteenuste kättesaamatusest tingitud passiivsus

Näitab, kui suur osa inimesi töötab osalise tööajaga või ei otsi tööd lapsehooldusteenuste puuduse tõttu. Osalise tööajaga töötavatel ja tööd mitteotsivatelt vastajatelt küsitakse osalise tööajaga töötamise ja töö mitteotsimise peamist põhjust. Kui peamine põhjus on lapsehoolduspuhkus või vajadus hoolitseda laste või hooldamist vajavate täiskasvanute eest, küsitakse, kas osalise tööajaga töötatakse või tööd ei otsita seetõttu, et sobiv lapsehooldusteenus puudub või on liiga kallis.

Andmeallikas: tööjõu-uuring

Ajavahemik: 2006–2013

Ühik: %

Meeste osatähtsus vanemahüvitise saajate hulgas

Näitab, kui suur osa vanemahüvitise saajatest on mehed. Samuti on võimalik arvutada keskmine vanemahüvitis soo järgi.

Andmeallikas: Maksu- ja Tolliameti maksukohustuslaste register

Ajavahemik: 2006–2013

Ühik: %, eurot

Meeste osatähtsus lapsehoolduspuhkuse kasutajate hulgas

Näitab, kui suur osa lapsehoolduspuhkuse kasutajatest on mehed.

Andmeallikas: Sotsiaalkindlustusameti riiklik pensionikindlustuse register

Ajavahemik: 2006–2013

Ühik: %

Isapuhkusel olnud meeste osatähtsus sündinud laste arvus

Näitab isapuhkust kasutanud meeste osatähtsust sama aasta elussündide arvus. Näitaja leidmiseks jagatakse ühe aasta jooksul isapuhkust kasutanud meeste arv samal aastal sündinud laste arvuga ja korrutatakse 100-ga.

Andmeallikas: Sotsiaalkindlustusameti riiklik pensionikindlustuse register

Ajavahemik: 2006–2013

Ühik: %

Meeste ja naiste ajakasutus

Näitab tasulise ja tasuta töö keskmist pikkust soo järgi ehk mitu minutit kulutavad mehed ja naised päevas tasuta ja tasulisele tööle.

Andmeallikas: ajakasutusuuring

Ajavahemik: 2000, 2010

Ühik: minutit

Vanemaks olemise mõju hõivelõhele

Näitab meeste ja naiste hõivelõhet alaealiste laste olemasolu järgi. Hõivelõhe on meeste ja naiste hõivemäärade erinevus protsendipunktides.

Andmeallikas: tööjõu-uuring

Ajavahemik: 2006–2013

Ühik: protsendipunkti

Meeste ja naiste loodud ettevõtete osatähtsus alustavate ettevõtete ning meeste ja naiste osatähtsus FIE-de hulgas

Näitab, kui suur on meeste ja naiste loodud ettevõtete osatähtsus alustavate ettevõtete hulgas. Samuti arvutatakse meeste ja naiste osatähtsus registreeritud FIE-de seas.

Andmeallikas: äriregister

Ajavahemik: 2006–2013

Ühik: %

ÜHENDANDMEBAASI ANDMETE VÕRDLUS TÖÖTASU STRUKTUURI UURINGU ANDMETEGA

Praegu kasutatakse Statistikaametis soolise palgalõhe arvutamiseks töötasu struktuuri uuringu (ingl Structure of Earnings Survey) andmeid. Uuringut tehakse iga nelja aasta tagant, viimane toimus 2014. aastal. Uuringu eesmärk on saada täpseid ja ühtlustatud andmeid EL-i elanike sissetulekute kohta. Sellega kogutakse infot töötajate töötasu, individuaalsete karakteristikute kohta (nt sugu, vanus, amet, tööstaaž, kõrgeim lõpetatud haridustase jne) ja tööandja karakteristikute (nt majanduslik aktiivsus, suurus ja ettevõtte asukoht) kohta^a. Töötasu struktuuri uuringu objektide arv üldkogumis oli 2014. aastal 80 500 ja valimis 10 900. Andmeid kogutakse veebipõhise elektroonilise andmete esitamise kanali eSTAT kaudu. Küsimustike laekumise jälgimiseks kasutatakse eSTATi. Küsimustikud on disainitud vastajale iseseisvaks täitmiseks kättesaadavad veebikeskkonnas, sisaldades juhiseid ja kontrole^b.

Vahepealsetel aastatel peavad liikmesriigid ise otsustama, kuidas nad palgalõhestatistikat teevad. Eestis kasutatakse selleks küsimustiku nr 1410 „Palk, lisa oktoobrikuu kohta“^c. Objektide (ettevõtete) arv üldkogumis oli 2013. aastal 81 361 ja objektide arv valimis 11 592. Riigi ja kohalike omavalitsuste asutusi ja vähemalt 50 töötajaga ettevõtteid vaadeldakse kõikselt, alla 50 töötajaga ettevõtetest tehakse stratifitseeritud lihtne juhuslik valik tegevusala järgi^d. Statistikaameti veebilehel on välja toodud palgastatistika üldkogum, valim, vastanute arv ja vastamismäär põhitegevusala (EMTAK 2008) järgi aastati^e. Valimisse sattunud ettevõtted täidavad töötajate kohta küsimustiku oktoobrikuu põhjal. Küsimustikust saadakse infot oktoobris töötasu saanud töötajate arvu, oktoobris töötatud tundide arvu ja tegelikult töötatud aja eest saadud tasu kohta soo järgi (vt tabel 14).

Oktoobrikuu lisa küsitakse valimisse sattunud ettevõtete käest andmeid töötajate palga ja töötundide kohta ettevõtte tasandil, mis tähendab, et kõikide ettevõtte töötajate andmed summeeritakse. Selle tulemusel määratakse kõikidele ühes ettevõttes töötavatele naistele ja meestele täpselt sama tunnipalk. Näiteks kui ettevõttes töötab kümme naist, kelle kuupalk ja töökoormus on erinev, summeeritakse nende kuupalgad ja töötatud tundide arv. Seejärel jagatakse summeeritud kuupalk töötatud tundide arvuga ja saadakse selle konkreetse ettevõtte naiste keskmine tunnipalk, mis omistatakse kõikidele selles ettevõttes töötavatele naistele. Samamoodi arvutatakse ka iga konkreetse ettevõtte meeste keskmine tunnipalk. Kuna mehi on kõrgepalgaliste seas märksa enam (nii ETU kui ka ESU 2012. aasta andmetel oli meeste osatähtsus kõrgeimas palgadetsiilis 80% ja naiste oma 20%, esimeses ehk madalaimas palgadetsiilis oli olukord täpselt vastupidine; vt täpsemalt kasutatavate andmeallikate ühilduvusanalüüsi^f), tõstab niisugune palgalõhe arvutamise meetodika meeste keskmist tunnipalka.

^a http://ec.europa.eu/eurostat/web/microdata/structure_of_earnings_survey

^b <http://www.stat.ee/esms-metaandmed?id=68932&code=21102>

^c http://www.stat.ee/public/files/aruandevormid/2015/21108141015_Aruandevorm_PDF-failina_et.pdf?t=1430294502

^d <http://www.stat.ee/68932?id=68932&code=21108>

^e Statistikaameti andmebaas, tabel PA5217: Palgastatistika üldkogum, valim, vastanud ja vastamisprotsent põhitegevusala (EMTAK 2008) järgi

^f <http://www.stat.ee/dokumendid/78174>

Tabel 14. Küsimustiku nr 1410 „Palk“ oktoobrikuu lisa

PALK, LISA OKTOOBRIKU KOHTA	
1	Tööandja koosseisu nimekirjas olnud ja töötasu saanud töötajate arv oktoobris – mehed
2	Töötatud tundide arv oktoobris – mehed (tuhandetes tundides kahe kümnendkohaga)
3	Tasu tegelikult töötatud aja eest – mehed
4	Tööandja koosseisu nimekirjas olnud ja töötasu saanud töötajate arv oktoobris – naised
5	Töötatud tundide arv oktoobris – naised (tuhandetes tundides kahe kümnendkohaga)
6	Tasu tegelikult töötatud aja eest – naised

EL-is kasutatakse soolise palgalõhe mõõtmiseks ja jälgimiseks korrigeerimata näitajat, mis tähendab, et palgaerinevuste arvutamisel ei ole arvestatud palgatöötaja individuaalseid karakteristikuid, mis võivad palgaerinevusi selgitada. Korrigeerimata palgalõhe näitab üldist naiste ja meeste vahelist ebavõrdsust, mida põhjustavad mitmesugused tegurid:

- naised ja mehed töötavad eri tegevus- ja ametialadel;
- naiste tööajakorraldus erineb meeste omast;
- nn naiste- ja meestetöid tasustatakse erinevalt.

Seega on sooline palgalõhe seotud mitme seadusest tuleneva, kuid ka sotsiaalse ja majandusliku teguriga, mis on tegelikult palju laiem teema kui lihtsalt võrdse töö ja palga idee.

Eurostati avaldatava soolise palgalõhe arvutamisel võetakse arvesse kõik palgatöötajad olenemata nende vanusest või töötatud tundide arvust. Samal ajal ei arvestata kõiki tegevusalasid ega alla 10 töötajaga ettevõtteid. Andmeid kogutakse majandustegevusalade kohta klassifikaatori EMTAK 2008 (Eesti majanduse tegevusalade klassifikaator) alusel ja selle järgi ei võeta arvesse palgatöötajaid, kes töötavad järgmistel tegevusaladel:

- põllumajandus, metsamajandus ja kalapüük (tegevusala tähtkood A);
- avalik haldus ja riigikaitse; kohustuslik sotsiaalkindlustus (tegevusala tähtkood O);
- kodumajapidamiste kui tööandjate tegevus; kodumajapidamiste oma tarbeks mõeldud eristamata kaupade tootmine ja teenuste osutamine (tegevusala tähtkood T);
- eksterritoriaalsete organisatsioonide ja üksuste tegevus (tegevusala tähtkood U).

Palgatöötajate palgana võetakse arvesse brutopalk, regulaarsed preemiad ja lisatasud ning ületunnitöö eest saadud tasu. Arvesse ei võeta ebaregulaarseid preemiaid. Töötundidena arvestatakse tööandja öeldud töötatud tundide arvu. See meetodika on ühesugune nii töötasu struktuuri uuringu aastatel kui ka uuringu vahepealsetel aastatel, kui soolist palgalõhet arvutatakse oktoobrikuu lisa põhjal.

Statistikaamet avaldab samuti soolist palgalõhet (tabel PA5335 „Sooline palgalõhe tegevusala (EMTAK 2008) järgi, oktoober“), mis tugineb küll Eurostati näitajaga samale andmeallikale, kuid mille arvutamise meetodikas on olulisi erinevusi. Eurostati andmebaasis avaldatav palgalõhe on suurem kui Statistikaameti andmebaasis avaldatav, nt 2013. aastal oli Eurostati andmetel palgalõhe Eestis 29,9%, kuid Statistikaameti andmetel 24,8%. Sellel on kaks peamist meetodilist põhjust. Esiteks võetakse Eurostati meetodika kohaselt palgalõhe arvutamisel arvesse ainult vähemalt 10 töötajaga ettevõtet, kuid Statistikaamet arvestab ka alla 10 töötajaga ettevõtteid. Kuna Eestis on palju just väikseid ehk alla 10 töötajaga ettevõtteid, kus meeste ja naiste palk ei pruugi kuigi palju erineda, ongi Statistikaameti avaldatud palgalõhe Eurostati omast väiksem. Teiseks võtab Eurostat tegevusalade klassifikaatori EMTAK 2008 järgi arvesse ainult tegevusalasid B–S, v.a O. Arvestamata jäävad seega

tegevusalad A (põllumajandus, metsamajandus ja kalapüük) ja O (avalik haldus ja riigikaitse; kohustuslik sotsiaalkindlustus). Statistikaamet aga arvestab kõiki tegevusalasid (A–S). Kuna põllumajanduse, metsamajanduse ja kalapüügi tegevusalal ning avaliku halduse ja riigikaitse ning kohustusliku sotsiaalkindlustuse tegevusalal on palgalõhe üldiselt väiksem kui teistel tegevusaladel, mõjutab nende tegevusalade väljajätmine palgalõhe suurust.

Eurostati avaldab soolist palgalõhet (ingl *gender pay gap in unadjusted form*) oma andmebaasis^a rubriigis „Population and social conditions“ → Labour market (labour)“ → „Earnings“ → „Gender pay gap in unadjusted form – NACE Rev. 2“. Statistikaamet avaldab soolist palgalõhet oma andmebaasis rubriigis „Majandus“ → „Palk ja tööjõukulu“ → „Palk“ → „Aastastatistika“ tabelis „PA5335: Sooline palgalõhe tegevusala (EMTAK 2008) järgi, oktoober“.

Tööjõu struktuuri uurigu tegemine iga nelja aasta tagant (2010, 2014 jne) on kohustuslik kõikidele Euroopa riikidele. Vahepealsetel aastatel otsustab aga iga riik ise, millist andmeallikat ta palgalõheandmete kogumisel kasutab. Kuna Eestis siiani kasutatav palgauuringu oktoobrikuu lisa on väga tömahukas, hakati projekti raames otsima võimalusi, kuidas vähendada andmeesitajate halduskoormust juba olemasolevate andmeallikate ja registriandmete kasutuselevõtuga. Projekti raames kokku kutsutud eksperdirühmaga otsustati 2013. aasta detsembris toimunud töötoas kasutada teiste andmeallikatena iga-aastaseid isiku-uuringuid (tööjõu-uuring ja sotsiaaluuring), millele on juurde lingitud täpsed palgaandmed Maksu- ja Tolliameti andmebaasist. Ühtlasi võimaldab palgalõhe arvutamise isiku-uuringute põhjal kaasata mitmesuguseid tausttunnuseid (nt haridustase, kõrgeima lõpetatud hariduse eriala, perekonnaseis, väikelaste olemasolu jms), mille kaudu palgalõhet ja selle dünaamikat süvitsi analüüsida.

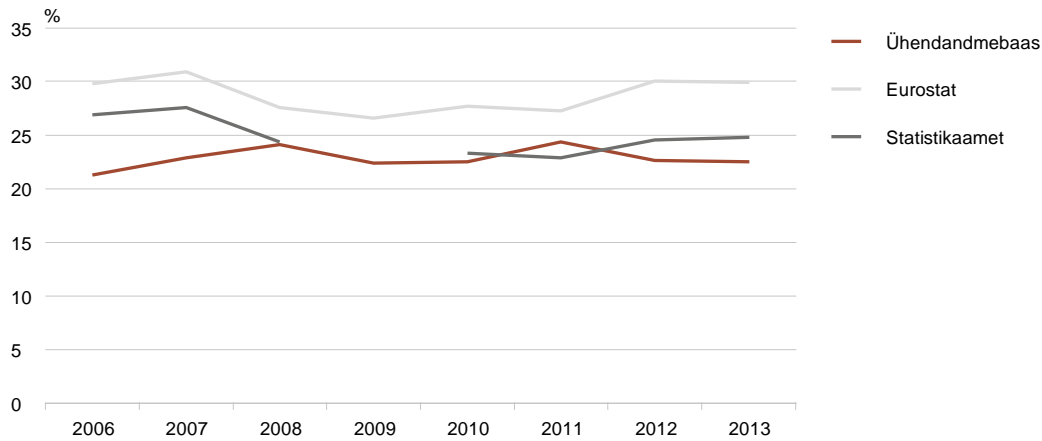
Töötasu struktuuri uuringu ja palgauuringu oktoobrikuu lisa ning palgaerinevuste projekti käigus loodud ühendandmebaasi põhjal arvatud sooline palgalõhe ei kattu üks ühele, sest on arvatud eri andmeallikate põhjal. Samuti on üks olulisi erinevusi see, et töötasu struktuuri uuring ja palgauuringu oktoobrikuu lisa on ettevõttepõhised, ühendandmebaas aga isikupõhine. See tähendab, et nimetatud uuringutes edastab ettevõtte oma töötajate andmed nende oktoobrikuu palga ja töötatud tundide kohta. Kuna isiku-uuringutes küsitakse vastaja käest tegelikult töötatud tundide arvu, annab selline lähenemisviis pildi sellest, mitu tundi vastajad tegelikult töötavad (mida tööandja ei pruugigi täpselt teada). Seega on võimalik palgalõhe arvutamiseks kasutada kaht viisi: ettevõttepõhist, mille puhul töötajate töötundide kohta vastab ettevõtte, ja isikupõhist, mille puhul vastab inimene ise, mitu tundi nädalas ta töötas. Ei saa öelda, et üks viis on õige ja teine vale, tegemist on eri meetoditega.

Palgauuringu oktoobrikuu lisa ja ühendandmebaasi erinevus on ka see, et kuna isiku-uuringutes küsitakse vastajalt erinevaid tausttunnuseid, on võimalik palgalõhet süvitsi analüüsida ja jälgida selle muutumist aja jooksul. Samal ajal on isiku-uuringute ühendandmebaasi puhul probleemiks see, et osal tegevusaladel on vähe vastajaid, mistõttu nende tegevusalade kohta usaldusväärseid andmeid ei saa.

Joonisel 4 on sooline palgalõhe Eurostati, Statistikaameti ja isiku-uuringute ühendandmebaasi järgi aastatel 2006–2013. Statistikaametil 2009. aasta kohta andmed puuduvad, sest sellel aastal töötasu struktuuri uuringut ega palgauuringu oktoobrikuu lisa ei tehtud. Jooniselt on näha, et Eurostati ja Statistikaameti näitajad liiguvad igal aastal samas suunas, kuid need on arvatud ka samade andmete põhjal. Ühendandmebaasi näitaja ühtib Statistikaameti avaldatuga enam kui Eurostati omaga, sest nende arvutamise metodika on sama (arvestatakse ka väikeettevõtteid ja kõiki tegevusalasid). Kõige rohkem erinevad palgalõhenäitajad aastatel 2006 ja 2007. Põhjus võib olla selles, et nendel aastatel oli väga paljudel isiku-uuringutes osalenutel isikukood puudu ja seda ei õnnestunud ka tagantjärele tuvastada. Seetõttu ei olnud võimalik neile lisada nende palgaandmeid Maksu- ja Tolliameti maksukohustuslaste registrist ega kaasata neid palgalõhe analüüsi.

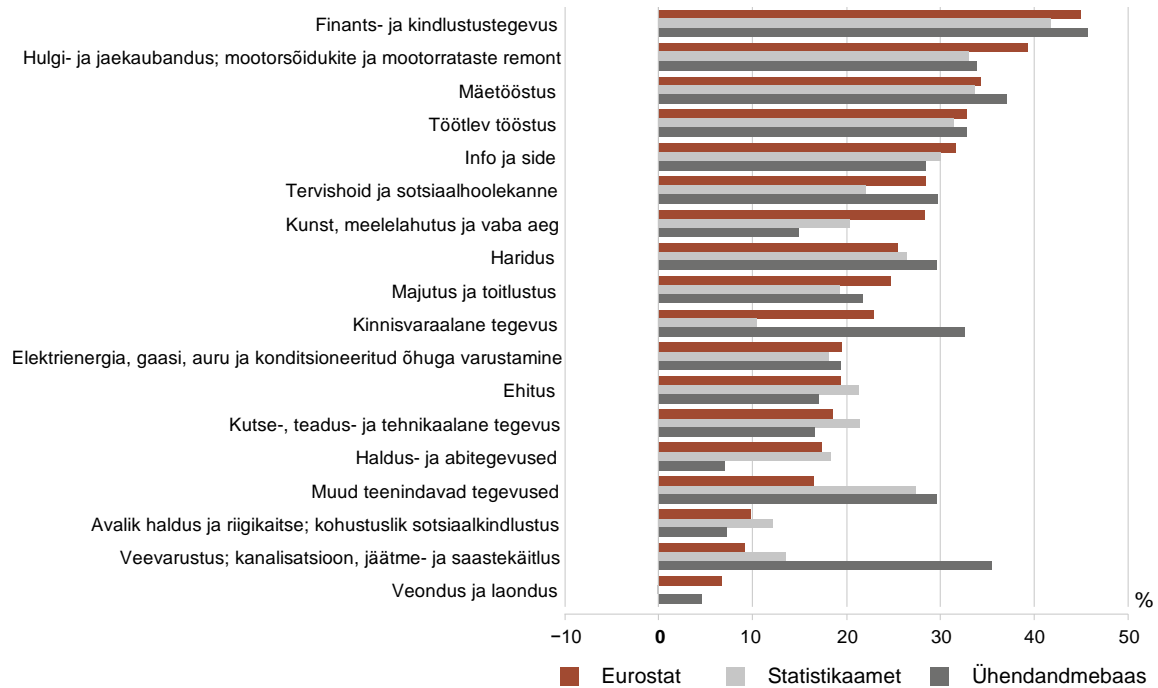
^a <http://ec.europa.eu/eurostat/data/database>

Joonis 4. Sooline palgalõhe Eurostati, Statistikaameti ja ühendandmebaasi järgi, 2006–2013



Joonisel 5 on palgalõhe tegevusala järgi Eurostati, Statistikaameti ja palgaerinevuste projekti käigus loodud ühendandmebaasi järgi (vt täpsemalt lisa 3).

Joonis 5. Sooline palgalõhe tegevusala järgi Eurostati, Statistikaameti ja ühendandmebaasi järgi, 2013



PALGAERINEVUSTE STATISTIKA STRATEEGIA

Statistikaametis praegu palgalõhe arvutamiseks kasutataval töötasu struktuuri uuringul, palgauuringu oktoobrikuu lisal ja palgaerinevuste projekti raames loodud ühendandmebaasil on nii positiivsed kui ka negatiivsed küljed.

Valimi suurus. Nii töötasu struktuuri uuringu kui ka palgauuringu oktoobrikuu lisa eelis on väga suur valimimaht: üle 10 000 ettevõtte ja töötasu struktuuri uuringus üle 100 000 isiku ning oktoobrikuu lisas üle 300 000 isiku. Isiku-uuringute ühendandmebaasis on palgalõhe arvutamise aluseks olevate isikute arv olenevalt aastast umbes 12 000 – 15 000.

Tegevusalad. Kuna isiku-uuringute ühendandmebaasi valim on võrreldes töötasu struktuuri uuringu omaga üsna väike, ei pruugi kõikide tegevusalade kohta saada kuigi usaldusväärseid andmeid. Näiteks on vähe vastajaid mäetööstuse ning elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamise tegevusalal.

Eri tausttunnused. Töötasu struktuuri uuringus on piiratud arv tausttunnuseid ja palgauuringu oktoobrikuu lisa järgi on palgalõhet võimalik vaadata vaid tegevusalati, isiku-uuringute ühendandmebaasis aga mitmesuguste tausttunnuste kaupa (vt loetelu peatükis „Palgalõhe ja seda mõjutavad näitajad“).

Töötatud tundide arv. Kuna isiku-uuringutes küsitakse töötaja enda käest, mitu tundi ta uuringunädalal töötas, on need andmed täpsemad kui ettevõtte öeldud töötundide arv töötasu struktuuri uuringus. Inimese enda öeldud tundide arv näitab, kui palju inimene tegelikult töötas. Siiski võib ka nende tundide puhul ette tulla probleem, nt kui inimene on aasta jooksul töötanud mitmes kohas. Seega läheb tõise tulu puhul arvesse küll terve aasta tööine tulu, kuid tundide arvu puhul ainult uuringunädalal töötatud tunnid, mis võivad aasta jooksul varieeruda.

Nimetatud positiivsete ja negatiivsete külgede tõttu on raske kindlaks teha, milline viis on palgalõhe arvutamiseks õige. Kuna ettevõttepõhine andmekogumine ei võimalda saada andmeid töötajate tausttunnuste kohta, on selleks praegu ainuke võimalus kasutada isiku-uuringuid, milles on rohkesti kõikvõimalikke tausttunnuseid, mille kaudu palgalõhet analüüsida.

Kuna Statistikaamet tegutseb üha enam selle nimel, et kasutada statistika tegemiseks registreid, on tulevikus alternatiivina võimalik hakata palgalõhet avaldama kõikse(te) registri(te) põhjal, mis annab vaieldamatult kõige täpsemad tulemused. Kahjuks ei ole praegu registrites veel vajalikul määral tausttunnuseid, kuid arendamisel on mitu registrit, mille põhjal oleks võimalik tulevikus palgalõhet koos mitmesuguste tausttunnustega avaldada.

Üks võimalik andmeallikas oleks töötamise register (TÖR), mis on Maksu- ja Tolliameti peetav töötamisega seotud infot koondav register^a. Kuna TÖR-i praegu veel arendatakse, ei saa palgalõhet kohe TÖR-i põhjal avaldada – see võib saada võimalikuks alles paari aasta pärast. Võimaluse korral lisatakse tulevikus TÖR-i veel tunnuseid või lingitakse seda teiste registritega, mistõttu oleks ka TÖR-i põhjal võimalik arvutada soolist palgalõhet mitmesuguste tausttunnuste kaupa.

Praegu on Statistikaamet TÖR-ist tellinud järgmised tunnused:

- töötaja isikukood
- töötaja eesnimi
- töötaja perekonnanimi
- töötaja sünniaeg
- töötamise liik, TÖR-sisene kood
- amet, TÖR-sisene kood
- tööd võimaldava isiku kood (äriregistri või isikukood)
- tööd võimaldava ettevõtte nimi, juhul kui tööd võimaldava isiku kood ei ole äriregistri kood või isikukood

^a <http://www.emta.ee/?id=35283>

- tegeliku tööandja äriregistri kood
- väljamakse tegija äriregistri kood
- tööd võimaldava isiku märkus
- töötamise asukoha kood
- töötamise algus
- töötamise lõpp
- töötamise lõpetamise alus, TÕR-sisene kood
- peatamise ID
- peatamise alus, TÕR-sisene kood
- peatamise perioodi algus
- peatamise perioodi lõpp
- esmane töötamise algus
- muutmise aeg
- sisestamise aeg
- töötamise kirje olek, TÕR-sisene kood

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
ISIK	Isiku unikaalne number uuringus				ETU: Aasta/Kvartal/Leibkond/Kysiti etav (Key2) ESU: Leibkond/Haru/Liige (Liikmenr)
STAT_KEY	Isiku unikaalne number Statistikaameti registris				Anonümiseeritud isikukood, stat_unit_id. Nendel isikutel, kellel uuringus isikukood puudus, on see tunnus puudu. Neile ei ole võimalik MTA andmeid linkida.
R_KUU	Uuringukuu		1–12		
UURING	Uuring	1 ETU 2 ESU			
I_SUGU	Isiku sugu	1 mees 2 naine			
SYNNIKUU	Isiku sünnikuu		1–12		
SYNNIAASTA	Isiku sünniaasta				
I_VANUS	Isiku vanus uuringuaasta alguse seisuga				Isiku vanus uuringuaasta alguse seisuga arvutatakse järgmiselt: uuringuaasta miinus sünniaasta miinus 1.
I_RAHVUS	Isiku rahvus	1 eestlane 2 venelane 3 muu 9 teadmata			
I_KODAK	Isiku kodakondsus	1 Eesti 2 Vene 3 muu 9 määramata kodakondsus			
I_PEREK_S	Isiku seaduslik perekonnaseis	1 vallaline, pole kunagi abielus olnud 2 abielus 3 lahutatud 4 lesk 9 teadmata			
I_PEREK_T	Isiku tegelik perekonnaseis	1 vallaline ja ei ole vabaabielus 2 abielus ja elab koos abikaasaga 3 vabaabielus 4 lahutatud ja pole uuesti abiellunud/kooselus 5 lesk ja pole uuesti abiellunud/kooselus 6 abielus ja elab seaduslikust abikaasast lahus, ei ole uuesti koosellu astunud 9 teadmata			
I_SYNNIRIIK	Isiku sünniriik	1 Eesti 2 muu riik 9 teadmata			
I_ELUK_MK	Elukoha maakonna kood	1 Tallinn 37 Harjumaa 39 Hiiumaa 44 Ida-Virumaa 49 Jõgevamaa 51 Järvamaa 57 Läänemaa 59 Lääne-Virumaa 65 Põlvamaa 67 Pärnumaa 70 Raplamaa			

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
		74 Saaremaa 78 Tartumaa 82 Valgamaa 84 Viljandimaa 86 Võrumaa 99 teadmata			
I_ELUK_TIH	Leibkonna elukoht asustustiheduse järgi	1 tihedalt asustatud 2 keskmise tihedusega 3 hõredalt asustatud 9 teadmata			Tihedusaste arvutatakse elukoha valla/linna koodi järgi. Reeglid selle kohta, millised vallad ja linnad millisesse asustustihedusega rühma kuuluvad, on ette antud. Alates 2012. aastast on jaotus järgmine: 1: Tallinn, Tartu, Narva; 2: Elva, Haapsalu, Jõgeva, Jõhvi, Keila, Kiviõli, Kuressaare, Maardu, Paide, Põlva, Pärnu, Rakvere, Saue, Sillamäe, Valga, Viljandi, Võru; 3: muu
I_ELUK_LM	Leibkonna elukoht: linn või maa	1 linn 2 maa 9 teadmata			
K_KEEL	Kodune keel	1 eesti 2 vene 3 muu 9 teadmata			
LAPS_0_3	Isiku leibkonnas elavate isiku kuni kolmeaastaste laste arv				
LAPS_4_6	Isiku leibkonnas elavate isiku 4–6-aastaste laste arv				
LAPS_7_17	Isiku leibkonnas elavate isiku 7–17-aastaste laste arv				
LAPS_18	Isiku leibkonnas elavate isiku vähemalt 18-aastaste laste arv				
LAPS_ALA	Isiku leibkonnas elavate isiku alaealiste laste arv (<18)				
LAPSI_LK	Isiku leibkonnas elavate isiku laste koguarv				
ISCED97M	Isiku kõrgeim haridustase ISCED97 järgi	1 tase 0 2 tase 1 3 tase 2a 4 tase 2c 5 tase 3a 6 tase 3b 7 tase 3c 8 tase 4b 9 tase 5a 10 tase 5b 11 tase 6 99 teadmata			Haridusküsimusi muudeti uuringutes aastal 2008. Seega olid aastatel 2006 ja 2007 küsimused ühtemoodi ja aastatel 2008–2013 teistmoodi. Ka see võib olla põhjus, miks andmetes on erinevusi.
I_ISCED6	Isiku kõrgeim haridustase ISCED97 kuue taseme järgi	0 eelharidus 1 põhihariduse alumine aste 2 põhihariduse ülemine aste			

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
		3 teise taseme haridus 4 kolmanda taseme eelne haridus 5 kolmanda taseme hariduse alumine aste 6 kolmanda taseme hariduse ülemine aste 9 teadmata			
I_ISCED3	Isiku kõrgeim haridustase ISCED97 kolme peamise taseme järgi	1 Esimese taseme haridus: alghariduseta, algharidusega, põhiharidusega, baashariduseta kutseharidus 2 Teise taseme haridus ja teise taseme järgne / kolmanda taseme eelne haridus: keskharidus, kutseõpe/ kutsekeskharidus/keskeriharidus põhihariduse baasil, kutsekeskharidus/kutseõpe keskhariduse baasil 3 Kolmanda taseme haridus: keskeriharidus keskhariduse baasil, kõrgharidus, magister, doktor 9 teadmata			
I_ERIALA_N	Isiku lõpetatud eriala nimetus	Klassifikaator: koolitusvalade klassifikaator			2006. ja 2007. aasta ESU-s on lõpetatud eriala kohta andmeid kogutud vaid kõige üldisemal tasemel, detailsem tase puudub.
I_ERIALA_K	Isiku lõpetatud eriala kood	Klassifikaator: koolitusvalade klassifikaator		999	
I_TOOLE	Isiku vanus, kui asus esimest korda pidevale tööle	0 pole kunagi töötanud 1–98 vanus täisaastates 99 teadmata		99	
T_SEIS_2	Sotsiaal-majanduslik seisund (töötamine = definitsiooni järgi, ülejäänud enesehinnanguline)	1 töötav 2 töötu 3 õppiv 4 pensionär 5 töövõimetuspensionär 6 ajateenija 7 kodune 8 muu mitteaktiivne			
I_TOOTAV	Isik on töötav	0 ei 1 jah			
I_AMET_N	Ametiala kirjeldus (klassifikaator – ametite klassifikaator 2008)	Klassifikaator: ametite klassifikaator 2008			
I_AMET_K	Ametiala neljakohaline kood (klassifikaator – ametite klassifikaator 2008)	Klassifikaator: ametite klassifikaator 2008		9999	
ISCO1	Isiku ametiala ühekohaline kood	1 juhid 2 tippspetsialistid 3 tehnikud ja keskastme spetsialistid 4 ametnikud 5 teenindus- ja müügitöötajad 6 põllumajanduse, metsanduse, jahinduse ja kalanduse oskustöötajad 7 oskustöötajad ja käsitöölised 8 seadme- ja masinaoperaatorid ja koostajad			

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
		9 lihttöölised 0 sõjaväelased 99 teadmata			
I_TEGEVUSAL A_K	Ettevõtte/asutuse üksuse kahekohaline tegevusala kood (klassifikaator: EMTAK 2008)	Klassifikaator: EMTAK 2008		999	
I_TEGEVUSAL A_N	Ettevõtte/asutuse üksuse tegevusala kood (klassifikaator: EMTAK 2008)	Klassifikaator: EMTAK 2008			
EMTAK1	Isiku tegevusala tähtkood	1 A põllumajandus, metsamajandus ja kalapüük 2 B mäetööstus 3 C töötlev tööstus 4 D elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine 5 E veevarustus; kanalisatsioon; jäätme- ja saastekäitlus 6 F ehitus 7 G hulgi- ja jaekaubandus; mootorsõidukite ja mootorrataste remont 8 H veondus ja laondus 9 I majutus ja toitlustus 10 J info ja side 11 K finants- ja kindlustustegevus 12 L kinnisvaraalaane tegevus 13 M kutse-, teadus- ja tehnikaalaane tegevus 14 N haldus- ja abitegevused 15 O avalik haldus ja riigikaitse; kohustuslik sotsiaalkindlustus 16 P haridus 17 Q tervishoid ja sotsiaalhoolekanne 18 R kunst, meelelahutus ja vaba aeg 19 S muud teenindavad tegevused 20 T kodumajapidamiste kui tööandjate tegevus; oma tarbeks kaupade tootmine ja teenuste osutamine 21 U eksterritoriaalsete organisatsioonide ja üksuste tegevus 99 teadmata			
I_V_TOOLE	Viimasele töökohale asumise aasta			9999	
T_AEG	Tööaja vorm	1 täistööaeg 2 osaline tööaeg 9 teadmata			
T_AEG_imp	Imputeeritud tööajavorm	1 täistööaeg 2 osaline tööaeg 9 teadmata			
t_aeg_i	Tööaja vorm on imputeeritud	0 ei 1 jah			
T_TUND	Tavaline nädalas töötatud tundide arv			999	ETU: 1) kui isik ütles oma tavalise nädalas töötatud

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
					tundide arvu, on see tavaline nädalas töötatud tundide arv; 2) kui tavaline nädalas töötatud tundide arv on teadmata, isik töötab täistööajaga ja tema eelmisel nädalal töötatud tundide arv on vahemikus 36–60, on tema tavaline nädalas töötatud tundide arv eelmisel nädalal töötatud tundide arv; 3) kui tavaline nädalas töötatud tundide arv on teadmata, isik töötab osalise tööajaga ja tema eelmisel nädalal töötatud tundide arv on vahemikus 11–35, on tema tavaline nädalas töötatud tundide arv eelmisel nädalal töötatud tundide arv.
T_TUND_imp	Imputeeritud tavaline nädalas töötatud tundide arv				
t_tund_i	Tavaline nädalas töötatud tundide arv on imputeeritud	0 ei 1 jah			
T_ASUK_LM	Töökoha asukoht	1 linn 2 maa 3 välismaa 9 teadmata			ESU-s seda küsimust ei ole.
T_ASUK_MK	Töökoha asukoha maakonna kood	1 Tallinn 37 Harjumaa 39 Hiiumaa 44 Ida-Virumaa 49 Jõgevamaa 51 Järvamaa 57 Läänemaa 59 Lääne-Virumaa 65 Põlvamaa 67 Pärnumaa 70 Raplamaa 74 Saaremaa 78 Tartumaa 82 Valgamaa 84 Viljandimaa 86 Võrumaa 97 välismaa 99 teadmata			ESU-s seda küsimust ei ole.
T_ARVGR1	Töötajate arv ettevõttes/asutuses/üksuses	1 1 kuni 10 2 11 kuni 19 3 20 kuni 49 4 vähemalt 50 9 teadmata			
T_ARV10	Töötajate arv kuni 10 töötajaga ettevõtetes	1 .. 10		99	
T_ARV2GR	Töötajate arv ettevõttes	1 kuni 10 2 vähemalt 11 9 teadmata			
T_SEISUND	Isiku tööalane seisund	1 palgatöötaja 2 palgatöötaja(te)ga ettevõtja, palgatööjõuga talupidaja 3 üksikettevõtja, palgatööjõuta talupidaja, vabakutseline 4 palgata töötaja pereettevõttes, talus			

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
		5 muu 9 teadmata			
T_STAATUS	Isiku staatus (ettevõtja/palgatöötaja)	10 ettevõtja 21 alaline palgatöötaja 22 ajutine palgatöötaja 99 teadmata			
I_K_TOO	Isikul on kõrvaltöö	1 jah 2 ei 9 teadmata			
I_PALGAT	Isik on palgatöötaja	0 ei 1 ja			
T_JUHT	Isikul on juhtimiskohustused	1 jah 2 ei 9 teadmata			
SEKTOR	Ettevõtte sektor	1 avalik sektor 2 erasektor 9 teadmata			2006. aasta ESU-s andmed puuduvad.
OMAND	Ettevõtte omandivorm	1 kodumaises omanduses 2 välisomanduses 3 muu: kodumaine välisosalusega 9 teadmata			2006. aasta ESU-s andmed puuduvad.
T_LEPING	Töölepingu tüüp	1 kirjalik 2 suuline 9 teadmata			
U_Kpalk_B	Uuringus öeldud tavaline brutokuupalk eurodes (ainult palgatöötajatel)				Täidetud ainult palgatöötajatel. 2006. aasta ESU-s küsimus puudus.
U_Kpalk_i	Uuringu brutokuupalk on imputeeritud	0 ei 1 jah			
Uuringukaal	Uuringusisene laiendustegur				ETU ja ESU isikukaalusid kasutatakse juhul, kui analüüsitakse ainult ühte uuringut.
Kaal	Andmebaasi laiendustegur				Uuringuaasta alguse seisuga 14-aastaste kaal on 0.
TT_S1_12_EA	Palk eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_12_E A	Juhatuse liikme tasu eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_12_E A	Lepingu alusel saadud tasu eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_12_EA	Muu tulu eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_01	Palk jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_01	Juhatuse liikme tasu				0 – ei ole saanud seda

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
	jaanuaris				tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_01	Lepingu alusel saadud tasu jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_01	Muu tulu jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_02	Palk veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_02	Juhatuse liikme tasu veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_02	Lepingu alusel saadud tasu veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_02	Muu tulu veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_03	Palk märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_03	Juhatuse liikme tasu märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_03	Lepingu alusel saadud tasu märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_03	Muu tulu märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_04	Palk aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_04	Juhatuse liikme tasu aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
TT_S15_04	Lepingu alusel saadud tasu aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_04	Muu tulu aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_05	Palk mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_05	Juhatuse liikme tasu mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_05	Lepingu alusel saadud tasu mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_05	Muu tulu mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_06	Palk juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_06	Juhatuse liikme tasu juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_06	Lepingu alusel saadud tasu juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_06	Muu tulu juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_07	Palk juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_07	Juhatuse liikme tasu juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_07	Lepingu alusel saadud tasu juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
MT_07	Muu tulu juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_08	Palk augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_08	Juhatuse liikme tasu augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_08	Lepingu alusel saadud tasu augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_08	Muu tulu augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_09	Palk septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_09	Juhatuse liikme tasu septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_09	Lepingu alusel saadud tasu septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_09	Muu tulu septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_10	Palk oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_10	Juhatuse liikme tasu oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_10	Lepingu alusel saadud tasu oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_10	Muu tulu oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub.

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
					Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_11	Palk novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_11	Juhatuse liikme tasu novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_11	Lepingu alusel saadud tasu novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_11	Muu tulu novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_12	Palk detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_12	Juhatuse liikme tasu detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_12	Lepingu alusel saadud tasu detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_12	Muu tulu detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
TT_S1_01_JA	Palk järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S11_01_JA	Juhatuse liikme tasu järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_S15_01_JA	Lepingu alusel saadud tasu järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_01_JA	Muu tulu järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
TT_12_EA	Tööne tulu kokku eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_01	Tööne tulu kokku jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_02	Tööne tulu kokku veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_03	Tööne tulu kokku märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_04	Tööne tulu kokku aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_05	Tööne tulu kokku mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_06	Tööne tulu kokku juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_07	Tööne tulu kokku juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_08	Tööne tulu kokku augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_09	Tööne tulu kokku septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_10	Tööne tulu kokku oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_11	Tööne tulu kokku novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_12	Tööne tulu kokku detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_01_JA	Tööne tulu kokku järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_aasta	Tööne tulu aastas kokku: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TT_kuude_arv	Nende kuude arv, kui töist tulu aastas saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
					MTA andmebaasis või isikukood puudub
TTS1_aasta	Palk aastas kokku: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TTS1_kuud	Nende kuude arv, kui palka aastas saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TTS11_aasta	Juhatuse liikme tasu aastas kokku: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TTS11_kuud	Nende kuude arv, kui juhatuse liikme tasu aastas saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TTS15_aasta	Lepingu alusel saadud tasu kokku aastas: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
TTS15_kuud	Nende kuude arv, kui lepingu alusel saadud tasu aastas saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
MT_aasta	Muu tulu kokku aastas: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
MT_kuude_arv	Nende kuude arv, kui muud tulu aastas saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_12_EA	Põhitööandja makstud palk eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_12_E A	Põhitööandja makstud juhatuse liikme tasu eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_12_E A	Põhitööandjalt lepingu alusel saadud tasu eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_12_EA	Põhitööandjalt saadud muu tulu eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_01	Põhitööandja makstud				0 – ei ole saanud seda

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
	palk jaanuaris				tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_01	Põhitööandja makstud juhatuse liikme tasu jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_01	Põhitööandjalt lepingu alusel saadud tasu jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_01	Põhitööandjalt saadud muu tulu jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_02	Põhitööandja makstud palk veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_02	Põhitööandja makstud juhatuse liikme tasu veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_02	Põhitööandjalt lepingu alusel saadud tasu veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_02	Põhitööandjalt saadud muu tulu veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_03	Põhitööandja makstud palk märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_03	Põhitööandja makstud juhatuse liikme tasu märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_03	Põhitööandjalt lepingu alusel saadud tasu märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_03	Põhitööandjalt saadud muu tulu märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_04	Põhitööandja makstud palk aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
PT_S11_04	Põhitööandja makstud juhatuse liikme tasu aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_04	Põhitööandjalt lepingu alusel saadud tasu aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_04	Põhitööandjalt saadud muu tulu aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_05	Põhitööandja makstud palk mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_05	Põhitööandja makstud juhatuse liikme tasu mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_05	Põhitööandjalt lepingu alusel saadud tasu mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_05	Põhitööandjalt saadud muu tulu mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_06	Põhitööandja makstud palk juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_06	Põhitööandja makstud juhatuse liikme tasu juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_06	Põhitööandjalt lepingu alusel saadud tasu juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_06	Põhitööandjalt saadud muu tulu juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_07	Põhitööandja makstud palk juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_07	Põhitööandja makstud juhatuse liikme tasu juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
PT_S15_07	Põhitööandjalt lepingu alusel saadud tasu juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_07	Põhitööandjalt saadud muu tulu juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_08	Põhitööandja makstud palk augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_08	Põhitööandja makstud juhatuse liikme tasu augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_08	Põhitööandjalt lepingu alusel saadud tasu augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_08	Põhitööandjalt saadud muu tulu augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_09	Põhitööandja makstud palk septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_09	Põhitööandja makstud juhatuse liikme tasu septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_09	Põhitööandjalt lepingu alusel saadud tasu septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_09	Põhitööandjalt saadud muu tulu septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_10	Põhitööandja makstud palk oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_10	Põhitööandja makstud juhatuse liikme tasu oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_10	Põhitööandjalt lepingu alusel saadud tasu oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
PT_MT_10	Põhitööandjalt saadud muu tulu oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_11	Põhitööandja makstud palk novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_11	Põhitööandja makstud juhatuse liikme tasu novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_11	Põhitööandjalt lepingu alusel saadud tasu novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_11	Põhitööandjalt saadud muu tulu novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_12	Põhitööandja makstud palk detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_12	Põhitööandja makstud juhatuse liikme tasu detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_12	Põhitööandjalt lepingu alusel saadud tasu detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_12	Põhitööandjalt saadud muu tulu detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_S1_01_JA	Põhitööandja makstud palk järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S11_01_JA	Põhitööandja makstud juhatuse liikme tasu järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_S15_01_JA	Põhitööandjalt lepingu alusel saadud tasu järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_01_JA	Põhitööandjalt saadud muu tulu järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub.

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
					Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_12_EA	Põhitööandjalt saadud teine tulu kokku eelmise aasta detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_01	Põhitööandjalt saadud teine tulu kokku jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_02	Põhitööandjalt saadud teine tulu kokku veebruaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_03	Põhitööandjalt saadud teine tulu kokku märtsis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_04	Põhitööandjalt saadud teine tulu kokku aprillis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_05	Põhitööandjalt saadud teine tulu kokku mais				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_06	Põhitööandja teine tulu kokku juunis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_07	Põhitööandjalt saadud teine tulu kokku juulis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_08	Põhitööandjalt saadud teine tulu kokku augustis				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_09	Põhitööandjalt saadud teine tulu kokku septembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_10	Põhitööandjalt saadud teine tulu kokku oktoobris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_11	Põhitööandjalt saadud teine tulu kokku novembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_12	Põhitööandjalt saadud teine tulu kokku detsembris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_01_JA	Põhitööandjalt saadud teine tulu kokku järgmise aasta jaanuaris				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_aasta	Põhitööandjalt saadud teine tulu aastas				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
	kokku: jaanuar kuni detsember				MTA andmebaasis või isikukood puudub
PT_kuude_arv	Nende kuude arv, kui põhitööandjalt töist tulu aastas saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PTS1_aasta	Põhitööandja makstud palk aastas kokku: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PTS1_kuud	Nende kuude arv, kui põhitööandjalt aastas palka saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PTS11_aasta	Põhitööandja makstud juhatuse liikme tasu aastas kokku: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PTS11_kuud	Nende kuude arv, kui põhitööandjalt juhatuse liikme tasu aastas saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PTS15_aasta	Põhitööandjalt lepingu alusel saadud tasu aastas kokku: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PTS15_kuud	Nende kuude arv, kui põhitööandjalt lepingu alusel aastas tasu saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
PT_MT_aasta	Põhitööandjalt saadud muu tulu aastas kokku: jaanuar kuni detsember				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub. Muudes tuludes ei ole arvestatud vanemahüvitist, sest MTA andmebaasis S19 = 0.
PT_MT_kuude_arv	Nende kuude arv, kui põhitööandjalt muud tulu aastas saadi				0 – ei ole saanud seda tululiiki; <i>missing</i> – ei sisaldu MTA andmebaasis või isikukood puudub
LISATUNNUSED					
T_ja_TT	Isik on uuringu järgi töötav ja on MTA järgi saanud töist tulu	0 ei 1 jah			
PT_ja_TT	Isik on uuringu järgi palgatöötaja ja on MTA järgi saanud töist tulu	0 ei 1 jah			
VAN_GR1	Vanuserühmad	1 kuni 24 2 25–34 3 35–44 4 45–54 5 55–64 6 65–74 7 75+			
Eriala_gr	Kõrgeima lõpetatud hariduse eriala	0 ei ole erialast haridust 1 õpetajakoolitus ja kasvatusteadus 2 humanitaar ja kunstid 3 sotsiaalteadused, äridus ja õigus			0 – töötab, aga erialast haridust ei ole

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
		4 loodus- ja täppisteadused 5 tehnika, tootmine ja ehitus 6 põllumajandus, metsandus ja kalandus, veterinaaria 7 tervis ja heaolu 8 teenindus 9 teadmata			
P_suhe	Isik on paarisuhtes	0 ei 1 jah			Ei ole paarisuhtes = vallaline või pole kunagi abielus olnud; lesk või lahutatud, kes pole uuesti abiellunud/kooselus; on paarisuhtes = on abielus/vabaabielus. Isikud, kelle tegelik perekonnaseis on teadmata, on liigitatud „ei ole paarisuhtes“ alla.
TT_ak	Tõine tulu: keskmiselt kuus kalendriaasta jooksul				Tõine tulu kalendriaastas kokku jagatud nende kuude arvuga, kui töist tulu saadi.
TT_uk	Tõine tulu: uuringukuul				Tõine tulu uuringukuul.
TT_ujk	Tõine tulu: uuringukuule järgneval kuul				Tõine tulu uuringukuule järgneval kuul.
TT_3k	Tõine tulu: uuringukuule eelneva, uuringukuu ja uuringukuule järgneva kuu keskmine				Tõine tulu uuringukuule eelneval, uuringukuul ja uuringukuule järgneval kuul kokku jagatud nende kuude arvuga, millal töist tulu saadi.
PT_ak	Põhitööandjalt saadud tõine tulu: keskmiselt kuus kalendriaasta jooksul				Tõine tulu kalendriaastas kokku jagatud nende kuude arvuga, millal töist tulu saadi.
PT_uk	Põhitööandjalt saadud tõine tulu: uuringukuul				Tõine tulu uuringukuul.
PT_ujk	Põhitööandjalt saadud tõine tulu: uuringukuule järgneval kuul				Tõine tulu uuringukuule järgneval kuul.
PT_3k	Põhitööandjalt saadud tõine tulu: uuringukuule eelneva, uuringukuu ja uuringukuule järgneva kuu keskmine				Tõine tulu uuringukuule eelneval, uuringukuul ja uuringukuule järgneval kuul kokku jagatud nende kuude arvuga, millal töist tulu saadi.
TT_tund	Kogu tõine tulu keskmiselt tunnis				Arvutamise valem: (kuukeskmine tõine tulu aastas * 12) / ((nädalas töötatud tunnid / 5) * tööpäevade arv aastas). Aasta tööpäevade arvuna on kasutatud 253, mis on aastate 2006–2014 keskmine. Keskmist on kasutatud, sest palk ei muutu, kui aastas on tööpäevi vähem või rohkem.
EMTAK10	EMTAKi 10 põhirühma	1 A põllumajandus, metsamajandus ja kalapüük 2 B–E tööstus 3 F ehitus 4 G–I kaubandus, transport,			Jaotus tugineb Eurostatis avaldatud NACE 10 jaotusele, kus: A – Agriculture, forestry and fishing

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
		majutus ja toitlustus 5 J info ja side 6 K finants- ja kindlustustegevus 7 L kinnisvaraalaane tegevus 8 M–N kutse-, teadus- ja tehnikaalaane tegevus, haldus- ja abitegevused 9 O–Q avalik haldus, haridus, tervishoid ja sotsiaahoolekanne 10 R–U kunst, meelelahutus ja vaba aeg, muud teenindavad tegevused, muu 99 teadmata			<i>B–E – Industry (except construction)</i> <i>F – Construction</i> <i>G–I – Wholesale and retail trade, transport, accommodation and food service activities</i> <i>J – Information and communication</i> <i>K – Financial and insurance activities</i> <i>L – Real estate activities</i> <i>M–N – Professional, scientific and technical activities; administrative and support service activities</i> <i>O–Q – Public administration, defence, education, human health and social work activities</i> <i>R–U – Arts, entertainment and recreation; other service activities; activities of household and extra-territorial organizations and bodies</i>
laps_EK	Eelkooliealiste laste arv (0–6)	0 ei ole eelkooliealisi lapsi 1 1 eelkooliealine laps 2 vähemalt 2 eelkooliealist last			
Juht	Juhid ISCO ja juhtimiskohustuse järgi	0 ei 1 jah			Kui ISCO1 kood = 1 ja isik on vastanud, et tal on juhtimiskohustusi.
Mees_LT	Isik on meeslihttöoline	0 ei 1 jah			Lähtudes TÜ analüüsi põhijäreldusest: meessoost palgatöoline, kes ametiklassifikaatori järgi on lihttöoline ja kelle ettevõtte tegevusala EMTAKi järgi on töötlev tööstus.
T_tund_s	Ettenähtud töötunnid			999	Leitud järgmiselt: osalise tööajaga ja töötab tavaliselt alla 35 tunni: tavaliselt töötatud tundide arv vastaja hinnangul; osalise tööajaga ja töötab tavaliselt vähemalt 35 tundi: 20 tundi (osalise tööajaga töötajate keskmine tundide arv); täistööajaga ja töötab hariduse valdkonnas (ameti kahekohaline kood = 23): 35 tundi; täistööajaga ja töötab mujal: 40 tundi.
TT_tund_s	Kogu tööine tulu keskmiselt tunnis ettenähtud tundide arvu järgi				Arvutamise valem: (kuukeskmine tööine tulu aastas * 12) / ((nädalas töötatud tunnid / 5) * tööpäevade arv aastas). Aasta tööpäevade arvuna on kasutatud 253, mis on aastate 2006–2014 keskmine. Keskmist on kasutatud, sest palk ei muutu, kui aastas on

Lisa 1. Ühendandmebaasi kirjeldus

Tunnuse nimetus	Tunnuse selgitus	Väärtuste selgitus	Väärtuse vahemik	Väärtused „Ei tea“ ja „Keeldub vastamast“	Märkus
Ettev	Isik on ettevõtja	0 ei 1 jah			tööpäevi vähem või rohkem. Isikud, kelle staatus on ettevõtja (T_STAATUS=10).
UP_tund	Brutotunnipalk uuringus öeldud palga alusel				
UP_tund_s	Tunnipalk uuringus öeldud brutopalga alusel (ettenähtud tundide arvu järgi)				
PT_tund	Põhitööandja makstud tunnipalk uuringus öeldud tundide arvu järgi				
PT_tund_s	Põhitööandja makstud tunnipalk ettenähtud tundide arvu järgi				

Lisa 2 (I). Osalise ja täistööajaga töötajate keskmise brutotunnipalga vahemikud tegevusala järgi, 2006–2009

(eurot)

	2006		2007		2008		2009									
	Osalise tööajaga –30% +30%	Täistöö- ajaga –30% +30%	Osalise tööajaga –30% +30%	Täistöö- ajaga –30% +30%	Osalise tööajaga –30% +30%	Täistöö- ajaga –30% +30%	Osalise tööajaga –30% +30%	Täistöö- ajaga –30% +30%								
Tegevusalade keskmine	1,95	3,62	2,52	4,68	2,35	4,36	3,05	5,67	2,78	5,16	3,51	6,53	2,88	5,35	3,40	6,32
Põllumajandus, metsamajandus ja kalapüük	1,69	3,13	2,04	3,78	2,03	3,77	2,47	4,58	2,41	4,47	2,84	5,28	2,15	3,98	2,65	4,92
Mäetööstus	2,02	3,76	2,89	5,37	2,43	4,52	3,50	6,51	2,89	5,36	4,03	7,49	3,01	5,60	3,98	7,40
Töötlev tööstus	1,71	3,17	2,35	4,37	2,05	3,81	2,85	5,30	2,43	4,51	3,29	6,10	2,92	5,43	3,22	5,97
Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine	1,81	3,37	2,93	5,44	2,18	4,05	3,55	6,59	2,58	4,80	4,09	7,59	2,94	5,47	4,40	8,18
Veevarustus; kanalisatsioon; jäätme- ja saastekäitlus	1,70	3,15	2,56	4,76	2,04	3,79	3,11	5,77	2,42	4,49	3,58	6,64	2,36	4,38	3,45	6,41
Ehitus	1,98	3,67	2,77	5,15	2,38	4,41	3,36	6,24	2,82	5,23	3,87	7,18	3,12	5,79	3,47	6,44
Hulgi- ja jaekaubandus; mootorsõidukite ja mootorrataste remont	1,59	2,96	2,44	4,53	1,92	3,56	2,96	5,49	2,27	4,22	3,40	6,32	2,39	4,43	3,19	5,92
Veondus ja laondus	1,75	3,25	2,47	4,58	2,11	3,91	2,99	5,55	2,50	4,64	3,44	6,39	2,65	4,92	3,28	6,09
Majutus ja toitlustus	1,48	2,74	1,70	3,15	1,77	3,29	2,06	3,82	2,10	3,91	2,37	4,40	2,15	3,99	2,32	4,31
Info ja side	2,50	4,64	3,95	7,34	3,01	5,58	4,79	8,90	3,56	6,62	5,52	10,25	3,63	6,74	5,49	10,20
Finants- ja kindlustustegevus	2,65	4,92	4,03	7,48	3,19	5,92	4,88	9,06	3,78	7,02	5,62	10,43	4,27	7,94	5,65	10,50
Kinnisvaraalane tegevus	1,67	3,10	1,97	3,66	2,01	3,73	2,39	4,43	2,38	4,42	2,75	5,10	2,59	4,82	2,77	5,14
Kutse-, teadus- ja tehnikaalane tegevus	2,18	4,05	2,93	5,44	2,62	4,87	3,55	6,60	3,11	5,78	4,09	7,60	3,39	6,30	3,93	7,29
Haldus- ja abitegevused	1,47	2,73	2,26	4,20	1,77	3,29	2,74	5,10	2,10	3,90	3,16	5,87	2,17	4,03	3,13	5,81
Avalik haldus ja riigikaitse; kohustuslik sotsiaalkindlustus	2,08	3,86	3,11	5,78	2,50	4,64	3,77	7,00	2,96	5,50	4,34	8,07	3,11	5,77	4,13	7,68
Haridus	2,22	4,13	2,23	4,15	2,67	4,96	2,71	5,02	3,17	5,88	3,12	5,79	3,13	5,81	3,10	5,77
Tervishoid ja sotsiaalhoolekanne	2,78	5,16	2,48	4,60	3,34	6,20	3,00	5,57	3,96	7,36	3,45	6,42	3,98	7,40	3,43	6,38
Kunst, meelelahutus ja vaba aeg	1,56	2,90	1,96	3,64	1,88	3,48	2,37	4,41	2,22	4,13	2,73	5,08	2,18	4,05	2,62	4,86
Muud teenindavad tegevused	1,41	2,63	1,87	3,48	1,70	3,16	2,27	4,22	2,02	3,74	2,62	4,86	2,03	3,77	2,51	4,65

Lisa 2 (II). Osalise ja täistööajaga töötajate keskmise brutotunnipalga vahemikud tegevusala järgi, 2010–2013

(eurot)

	2010		2011		2012		2013									
	Osalise tööajaga		Täistööajaga		Osalise tööajaga		Täistööajaga									
	-30%	+30%	-30%	+30%	-30%	+30%	-30%	+30%								
Tegevusalade keskmine	2,82	5,25	3,39	6,30	2,91	5,40	3,58	6,64	2,98	5,54	3,79	7,05	3,20	5,94	4,07	7,57
Põllumajandus, metsamajandus ja kalapüük	2,15	3,99	2,79	5,18	2,32	4,30	2,80	5,20	2,42	4,50	3,07	5,69	2,58	4,78	3,57	6,63
Mäetööstus	2,90	5,38	4,20	7,80	3,47	6,44	4,54	8,44	3,56	6,62	4,70	8,72	3,84	7,14	5,19	9,65
Töötlev tööstus	2,66	4,95	3,23	6,01	2,81	5,23	3,39	6,29	2,84	5,28	3,69	6,85	3,07	5,69	3,97	7,37
Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine	2,89	5,37	4,51	8,38	3,33	6,18	4,84	9,00	3,49	6,47	5,40	10,04	3,69	6,85	5,84	10,84
Veevarustus; kanalisatsioon; jäätme- ja saastekäitlus	2,18	4,05	3,41	6,32	1,93	3,58	3,49	6,49	2,07	3,85	3,84	7,14	2,25	4,19	4,11	7,63
Ehitus	2,74	5,09	3,49	6,49	3,22	5,98	3,62	6,72	3,19	5,93	4,00	7,42	3,18	5,90	4,28	7,94
Hulgi- ja jaekaubandus; mootorsõidukite ja mootorrataste remont	2,37	4,41	3,10	5,75	2,41	4,47	3,42	6,34	2,52	4,68	3,60	6,68	2,66	4,94	3,77	7,01
Veondus ja laondus	2,72	5,05	3,33	6,19	2,32	4,30	3,40	6,31	2,56	4,75	3,45	6,41	2,79	5,17	3,86	7,16
Majutus ja toitlustus	2,01	3,74	2,19	4,07	2,16	4,00	2,32	4,32	2,23	4,15	2,47	4,59	2,21	4,11	2,63	4,88
Info ja side	3,71	6,90	5,49	10,19	3,82	7,10	5,97	11,09	3,84	7,14	6,19	11,49	4,38	8,14	6,67	12,39
Finants- ja kindlustustegevus	4,77	8,87	5,50	10,21	5,42	10,06	5,71	10,61	5,82	10,80	5,90	10,96	6,18	11,48	6,44	11,96
Kinnisvaraalaane tegevus	2,09	3,87	2,35	4,36	2,33	4,33	2,67	4,95	2,39	4,43	2,84	5,27	2,54	4,72	2,98	5,54
Kutse-, teadus- ja tehnikaalaane tegevus	3,80	7,05	4,16	7,73	3,54	6,57	4,38	8,13	3,93	7,29	4,87	9,04	3,86	7,18	4,85	9,01
Haldus- ja abitegevused	2,19	4,07	3,16	5,87	2,16	4,00	3,30	6,12	2,11	3,91	3,44	6,38	2,37	4,39	3,74	6,94
Avalik haldus ja riigikaitse; kohustuslik sotsiaalkindlustus	3,14	5,83	4,05	7,52	3,19	5,92	4,15	7,71	3,49	6,49	4,34	8,06	3,76	6,98	4,75	8,81
Haridus	3,06	5,68	3,09	5,74	3,10	5,76	3,16	5,86	3,19	5,93	3,26	6,06	3,51	6,51	3,53	6,55
Tervishoid ja sotsiaalhoolekanne	3,82	7,10	3,28	6,09	4,07	7,55	3,44	6,38	4,07	7,55	3,63	6,73	4,44	8,24	3,90	7,24
Kunst, meelelahutus ja vaba aeg	2,57	4,78	2,68	4,98	2,26	4,20	2,91	5,40	2,28	4,23	3,21	5,95	2,44	4,52	3,07	5,71
Muud teenindavad tegevused	1,84	3,42	2,28	4,23	1,86	3,45	2,06	3,82	1,79	3,33	2,21	4,10	1,82	3,38	2,37	4,39

Lisa 3. Sooline palgalõhe Eurostati, Statistikaameti ja ühendandmebaasi järgi, 2013

(protsenti)

Majandustegevusala	Eurostat	Statistikaamet	Ühendandmebaas
Tegevusalad kokku: A–S	ei avalda	24,8	22,6
Tegevusalad B–S, v.a O	29,9 ^a	ei avalda	25,7
Tegevusalad B–S, sh O	28,2	ei avalda	23,5
A. Põllumajandus, metsamajandus ja kalapüük	ei avalda	12,7	20,7
B. Mäetööstus	34,3	33,7	37,0
C. Töötlev tööstus	32,8	31,5	32,8
D. Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine	19,6	18,1	19,4
E. Veevarustus; kanalisatsioon, jäätme- ja saastekäitlus	9,2	13,6	35,5
F. Ehitus	19,4	21,3	17,1
G. Hulgi- ja jaekaubandus; mootorsõidukite ja mootorrattaste remont	39,4	33,0	33,9
H. Veondus ja laondus	6,7	–0,2	4,5
I. Majutus ja toitlustus	24,7	19,3	21,8
J. Info ja side	31,6	30,1	28,5
K. Finants- ja kindlustustegevus	44,9	41,8	45,7
L. Kinnisvaraala tegevus	22,9	10,5	32,6
M. Kutse-, teadus- ja tehnikaala tegevus	18,5	21,4	16,6
N. Haldus- ja abitegevused	17,4	18,4	7,0
O. Avalik haldus ja riigikaitse; kohustuslik sotsiaalkindlustus	9,8	12,1	7,2
P. Haridus	25,5	26,4	29,6
Q. Tervishoid ja sotsiaalhoolekanne	28,5	22,0	29,8
R. Kunst, meelelahutus ja vaba aeg	28,4	20,3	14,9
S. Muud teenindavad tegevused	16,5	27,4	29,6

^a Euroopa Liidus kasutatava palgaerinevuste näitaja mõõtmiseks ja jälgimiseks on arvesse võetud ainult tegevusalad B–S, tegevusala O on välja jäetud. Samuti ei ole arvestatud alla 10 töötajaga ettevõtteid.

Allikad: Statistikaamet, Eurostat^b

^b Statistikaameti näitajad: Statistikaameti andmebaasi tabel „PA5335: Sooline palgalõhe tegevusala (EMTAK 2008) järgi, oktoober“; Eurostati näitajad: Eurostati andmebaasi tabel „Gender pay gap in unadjusted form in % – NACE Rev. 2 (structure of earnings survey methodology)“ (earn_gr_gpgr2).

INTRODUCTION

Gender equality is one of the founding principles of the European Union (EU)^a. The principle of equal pay for equal work has been present in EU treaties since 1957 and it is also incorporated in Directive 2006/54/EC^b on equal treatment between women and men in employment and occupation. There are several causes of the gender pay gap: uneven gender distribution in education and on the labour market, difficulties with reconciliation of work and family life, outdated staff practices, etc. Currently, data for calculating gender pay gap are collected with the Structure of Earnings Survey, which is conducted every four years. In the intermediate years, gender pay gap is calculated based on the questionnaire No. 1410 "Palk, lisa oktoobrikuu kohta" ("Wages and salaries, annex for October")^c. As data collection is company-based, it is not possible to collect background variables about the employee which would enable in-depth analysis of pay gap dynamics. In order to assess the causes of the pay gap, it is necessary to regularly collect various background variables about the employees, e.g. their level of education, field of education and training, occupation and economic activity, working time format, marital status, presence of (small) children, etc.

The first chapter of this present methodological report provides a brief introduction to the project "Increased availability of gender pay gap statistics" and its main objectives. Chapter two gives a short overview of the methodology and data sources used for calculating the gender pay gap up to now. Chapter three explains the creation of the combined database. Chapter four reviews additional indicators that can be published along with the gender pay gap. Chapter five compares the data sources used so far to calculate the gender pay gap to the new combined database, which was created in the framework of the project, with regard to their methodological differences and similarities in their data. The last chapter – chapter six – focuses on the strategy of pay gap statistics. At the end of the publication there is also an overview of all the variables published in the combined database.

^a http://europa.eu/rapid/press-release_IP-14-190_en.htm

^b <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0054&qid=1442992515398&from=EN>

^c http://www.stat.ee/public/files/aruandevormid/2015/21108141015_Questionnaire_as_a_PDF_file_en.pdf?t=1442993628

PROJECT INTRODUCTION

Currently, data for calculating the gender pay gap are collected with the Structure of Earnings Survey which is obligatory for all EU Member States. The Structure of Earnings Survey is conducted in the European Union every four years. As of 2008, Estonia no longer conducts that survey annually either (it was not conducted in 2009). However, considering that survey data are collected through companies, very few background variables are collected about the person receiving the wages (salaries).

At the same time, there are other data sources used annually to collect information about earnings, such as individual surveys and registers. Introduction of already existing data sources would significantly reduce the administrative burden of respondents. Additionally, individual surveys enable in-depth analysis of the gender pay gap, as they comprise many different background variables (e.g. employee's level of education, field of education and training, occupation and economic activity, working time format, marital status, presence of (small) children, etc.).

In the course of the project "Increased availability of gender pay gap statistics", already existing data sources (surveys, registers) were mapped to find the best data sources for monitoring the gender pay gap. In the workshop held in December 2013, it was decided with the expert group convened under the project, that, based on previous analyses, the most suitable solution would be to combine the samples from the individual surveys, the Estonian Labour Force Survey and the Estonian Social Survey (for the purpose of increasing the sample), and link to them exact wage data of the Estonian Tax and Customs Board. Then an analysis was conducted on whether calculation of the pay gap based on a combined database created by using these surveys would ultimately be easier, quicker and cheaper.

Thus, the object of the project was to develop a new gender pay gap calculating methodology which can be used to make proposals for methodological improvements and the utilisation of data sources behind the calculation of the pay gap. This would also reduce the administrative burden of respondents (companies) and Statistics Estonia.

Using new data sources will allow calculating and publishing the gender pay gap:

- in a timely manner, meaning that the data of the previous year can be published within the first half of the following year at the latest (for example, the gender pay gap of 2015 is published within the first half of 2016);
- with relevant indicators that describe the background of the gender pay gap and allow policy-makers to rely on knowledge-based analysis.

Project activities included convening an expert group and organising regular workshops, where the experts of the working group and representatives of stakeholders could participate in discussions on data sources, database variables, additional indicators, and methodological decisions. For more detailed information about the decisions made and the subjects discussed in workshops, see project homepage^a.

The project is funded via the Norwegian support programme "Mainstreaming Gender Equality and Promoting Work-Life Balance" for 2009–2014. Updated information about the project's activities is available on the homepage of Statistics Estonia under "Economy" > "Wages and salaries and labour costs".

^a <http://www.stat.ee/palgaerinevuste-statistika-parem-kattesaadavus> (only in Estonian)

CALCULATING THE GENDER PAY GAP

The Statistical Office of the EU (Eurostat) commenced the harmonisation of calculation of the gender pay gap in 2005. Until 2005, the sources used for the calculation of the gender pay gap varied by country – administrative data, the Labour Force Survey, EU Statistics on Income and Living Conditions, or other nationwide survey. This brought about differences in definitions, total population, sample size, etc., which affected the quality of the gender pay gap indicator and its comparability across Member States. Therefore, it became necessary to start using a common data source and develop a common methodology for calculating the gender pay gap.

In the EU, the indicator used to monitor the gender pay gap is the gender pay gap in unadjusted form, published annually by Eurostat for all Member States. The gender pay gap indicates the difference between the wages of women and men. Calculation of this indicator is based on hourly gross wages and it shows the percentage by which females' hourly gross wages are smaller than those of males. The formula is the following:

$$\frac{\text{average hourly gross wages of males} - \text{average hourly gross wages of females}}{\text{average hourly gross wages of males}} * 100$$

Gender pay gap data which are published by Eurostat are collected with the Structure of Earnings Survey (SES), which is mandatory for Member States to conduct every four years. The most recent Structure of Earnings Survey was carried out in 2014. As pay gap data is published annually, some countries conduct the survey annually as well, while others use alternative data sources in the intermediate years. In those years, in Estonia, the pay gap is calculated by using data from the questionnaire "Wages and salaries, annex for October"^a. In October each year (except for in the year of Structure of Earnings Survey), the enterprises included in the sample will fill out a module in the questionnaire, the data of which is used to calculate the gender pay gap.

^a <http://www.stat.ee/65596>

ESTABLISHING THE COMBINED DATABASE

One key objective of the project “Increased availability of gender pay gap statistics” is to create a combined database that would contain not only data on wages, but also other significant variables necessary for a complex analysis of the gender pay gap. The variables, in turn, help to understand the essence of this phenomenon and provide an explanation for its causes. On the one hand, the presence of a common and comprehensive database allows performing in-depth analyses that provide a direction and input for developing policies, and on the other hand it gives an opportunity to observe changes in society.

The process of mapping potential data sources to base the new combined database on and discussion about the variables to be included in the database began immediately after the start of the project. In December 2013, a meeting was held with the specialists of the working group convened under the project. In the meeting, key decisions were made concerning the data sources to be used for creating the combined database and variables to be published in the database. Users could also participate in the discussion – it was possible to submit proposals on the website of Statistics Estonia. The materials of the workshop are accessible on the project’s homepage^a.

Main decisions made upon the establishment of the combined database:

- the database is established on the basis of the Estonian Labour Force Survey and the Estonian Social Survey;
- due to the rotation of the Labour Force Survey, one person may participate in the survey up to twice a year. It was decided that all persons will be treated as first-time participants in the survey, i.e. the rotation will not be taken into account;
- database will take into account all persons who, based on these surveys, are wage employees;
- the combined database will be linked with data on wages i.e. earned income collected by the Estonian Tax and Customs Board (ETCB).

The group also agreed on the variables to be included in the new combined database that would allow further in-depth analyses and the calculation of other significant indicators to be published along with the pay gap on a regular basis. The final combined database contains the following variables (for details, see Annex 1):

- sex;
- month of birth;
- year of birth;
- age as at the beginning of the survey year;
- ethnic nationality;
- citizenship;
- legal marital status;
- actual marital status;
- country of birth;
- place of residence;
- household’s place of residence
- household’s place of residence by population density;
- main language spoken in the household;
- number of children in the household of the person;
- ages of the children in the household of the person;
- highest level of education attained;

^a <http://www.stat.ee/75328> (only in Estonian)

- *field of education attained;*
- *age upon the first commencement of continuous employment;*
- *year of commencement of work in the last job;*
- *socioeconomic status;*
- *occupation;*
- *economic activity;*
- *working time format;*
- *number of hours usually worked per week;*
- *workplace location;*
- *number of employees in the company;*
- *employment status;*
- *status (entrepreneur or wage employee);*
- *presence of off-hour job;*
- *presence of managerial responsibilities;*
- *employment contract type;*
- *membership in trade union;*
- *economic sector of the company;*
- *form of ownership of the company;*
- *size of the company;*
- *usual gross monthly wages stated in the survey;*
- *earned income based on the data of the ETCB.*

After creating the database it was important to establish the list of indicators that would be published regularly along with the pay gap. However, it is impossible to find all the indicators related to the gender pay gap based on the combined database alone. This is due to the fact that the data necessary are not collected with surveys or because these social processes concern such a small group of people that it is impossible to gather data about them with a survey. Therefore, the next step was to decide on the indicators that would help to understand the reasons behind the occurrence and persistence of a gender pay gap and should certainly be published together with the pay gap.

In the workshop held in December 2013, it was decided that the indicators will be based on the recommended indicators published in the publication “Sooline palgalõhe Eestis. Artiklite kogumik” (“Gender pay gap in Estonia. Collection of articles”) of the Ministry of Social Affairs. These indicators allow monitoring the differences in payment for work and the role of (structural) factors therein^a.

At the expert working group meeting held in March 2015 it was decided to publish pay gap time series (2006–2013) along with the following indicators:

- *average gender pay gap;*
- *gender pay gap by the following variables:*
 - *age*
 - *ethnic nationality*
 - *place of residence*
 - *level of education*
 - *field of highest level of education attained*
 - *occupation*
 - *economic activity*

^a https://www.sm.ee/sites/default/files/content-editors/Ministeerium_kontaktid/Valjaanded/toimetised_20112.pdf (only in Estonian)

- *working time format*
- *presence of managerial responsibilities*
- *type of employer*
- *size of employer*
- *marital status*
- *presence of children*
- economic activity- and occupation-based segregation in the labour market;
- share of men and women among managers;
- share of men and women among entrepreneurs;
- share of companies founded by men and women among new companies;
- share of men and women among registered sole proprietorships;
- share of part-time workers among female and male wage employees;
- passivity caused by unavailability of child care services;
- share of men among the recipients of parental benefit;
- average monthly amount of parental benefit of women and men;
- share of men among the users of parental leave;
- use of paternity leave;
- time use by sex;
- impact of parenthood on employment.

Indicators that are not published on a regular basis are the explained and unexplained pay gap indicators (for 2011 and 2012; next recommended year is for 2015)

The materials and additional information about the decisions made at the meeting of experts' working group are accessible on the project website^a.

Basis for establishing the combined database

Creating a combined database starts with combining the variables of the Estonian Labour Force Survey (ELFS) and the Estonian Social Survey (ESS). At first, the methodology of collecting variables of both surveys is analysed and, in case of discrepancies, the original variables are re-coded so as to make them comparable. Then it is possible to merge the databases.

Based on a unique identifying variable developed by Statistics Estonia, the combined database will be supplemented with the data from the register of taxable persons of the Estonian Tax and Customs Board (ETCB). However, here it is necessary to take into account that, in the ETCB database, one person does not equal one entry; one entry includes the transactions made to one person by one company in one month. Therefore, in order to add ETCB data to the data of persons who participated in individual surveys, the ETCB database must be rearranged to an individual level in a way that there would only be one entry per person.

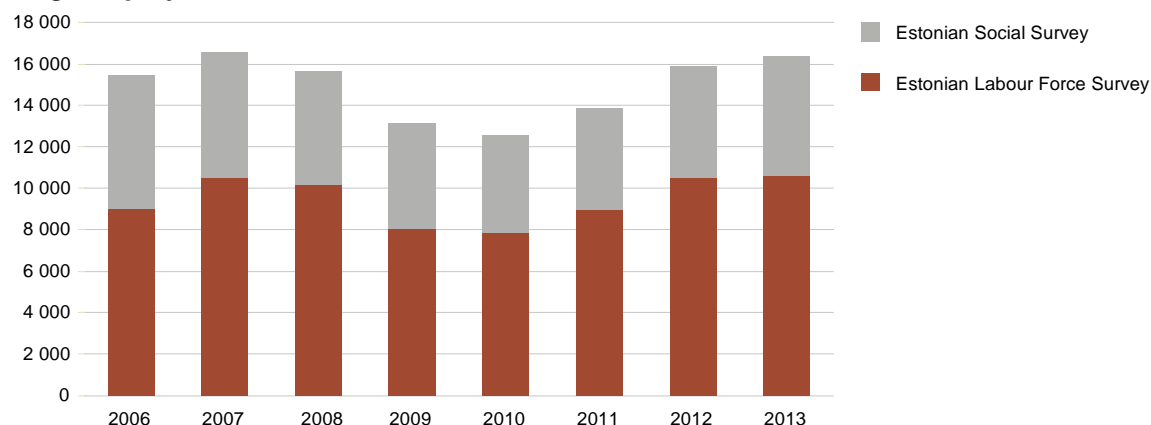
A new expansion factor (weight) will be determined for the new combined database, allowing generalisation of the results for the entire population.

As the annual sample of the ELFS exceeds that of the ESS, there are more persons that participated in the ELFS in the combined database than there are those who participated in the ESS. Figure 1 shows the number of people participating in the two surveys in 2006–2013. It also separately refers to the number of wage employees because this is the basis for calculating the gender pay gap.

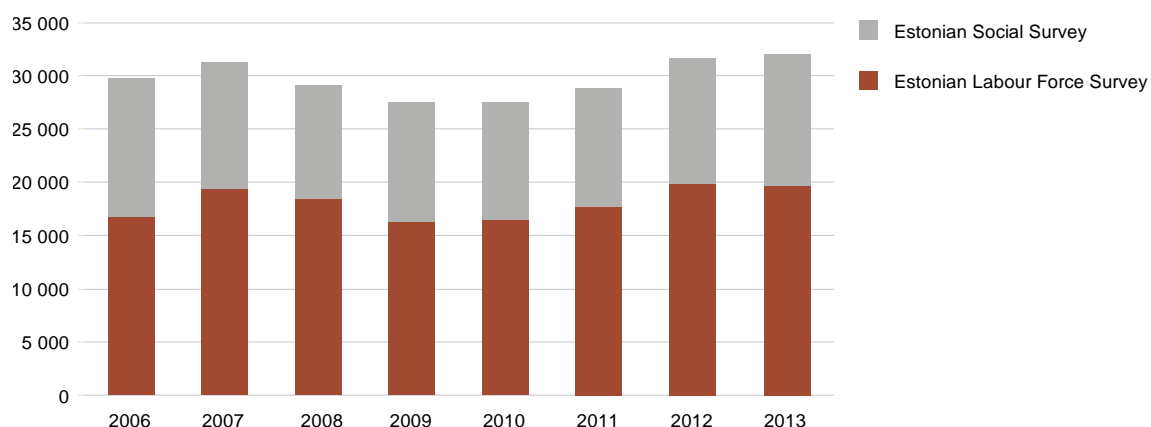
^a <http://www.stat.ee/183300> (only in Estonian)

Figure 1. Number of persons and wage employees in the combined database by survey, 2006–2013

Wage employees



Persons



The analysis below describes the creation of the combined database, the problems that occurred along with the relevant solutions, and compatibility of the variables that are based on different data sources.

Combining the data from the Estonian Labour Force Survey and the Estonian Social Survey

Re-coding variables

The prerequisite for combining the variables of two surveys is that the variables of both surveys are collected in the same manner, i.e. the questions have the same meaning and provide the same answer options. A problem occurs in creating the combined database if some of the variables are collected by only one survey, variables are collected differently (e.g. different reply options) or a question about a variable is asked from dissimilar groups. In that case, the variable needs to be re-coded.

Variables collected by only one survey

When combining the variables of the ELFS and the ESS, a problem occurs in case of the variable of the location of the workplace which is only included in the ELFS. This variable is not present for those participating in the ESS.

Variable collected differently in the surveys

Another problem is that the method for collecting variables differs in the ELFS and the ESS.

The first such variable is the working time format, i.e. whether the person works part time or full time. This information should be available for all persons who are employed at the survey moment. In the ELFS, the respondents provide their subjective assessment of whether they work full time or part time. But the ESS only asks for an assessment of the respondent's socioeconomic status, including working time format. The possible answers are the following:

- *full-time wage employee;*
- *part-time wage employee;*
- *full-time entrepreneur-employer, self-employed person, farmer without paid employees, freelancer, unpaid family worker (in a farm);*
- *part-time entrepreneur-employer, self-employed person, farmer without paid employees, freelancer, unpaid family worker (in a farm);*
- *unemployed;*
- *(university) student, unpaid intern;*
- *in early retirement or old-age pensioner;*
- *pensioner receiving pension for incapacity for work;*
- *homemaker, person on parental leave;*
- *conscript;*
- *other inactive.*

In order to ensure that the variable collected in both surveys would rely on the respondent's self-assessment, it is necessary to use subjective assessment in the ESS as well. However, there is a problem with persons with two statuses (for example, when a person is employed but also retired). These persons might not consider employment as their main status and thus it remains unknown whether the person actually worked full-time or part-time. Thus, for the time being, the working time format remains occasionally unknown in the ESS, but it is imputed in the course of further processing.

Both surveys ask the number of hours usually worked per week which can also be used for assessing whether a person worked full time or part time. For that purpose, it is first necessary to check how well the self-assessment of the respondent matches the working time format based on hours usually worked. According to international definition, persons working part time are persons who work less than 35 hours per week. Those who work at least 35 hours are full-time employees.

A comparison between the two variables shows that self-assessment and number of hours usually worked per week coincide in the case of more than 90% of people, meaning that for people without the variable of self-assessed working time format, it can be derived from the number of hours usually worked per week. A more detailed analysis about the establishment of the variable of working time format is available in the compatibility analysis of used data sources^a.

Table 1 shows the total number of entries and employed persons in the combined database of the ELFS and ESS. The table also shows the number of people for whom it was possible to determine the working time format, and the number of people for whom it was impossible. In the latter case, net wages were calculated by statistical methods through other variables, i.e. net wages were imputed. A more detailed description of imputation methodology is available in the chapter "Imputation of working time format".

^a <http://www.stat.ee/dokumendid/78174> (only in Estonian)

Table 1. Presence of working time format variable in the combined database, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Entries	29,793	31,284	29,131	27,466	27,604	28,734	31,616	31,971
Employed persons	16,803	18,092	16,933	14,340	13,690	15,161	17,443	17,915
Working time format present	16,803	18,090	16,932	14,339	13,686	15,155	17,434	17,901
Working time format not present	0	2	1	1	4	6	9	14

Another variable that is collected differently in the ELFS and the ESS is the number of hours usually worked per week which makes it impossible to combine these variables automatically. Both surveys use the same question: “How many hours per week do you usually work at this job?”

The ELFS also allows the response “hard to say, hours worked vary considerably from week to week or from month to month”. Furthermore, the ELFS features a question about the number of hours worked during the survey week (by day) which is not present in the ESS.

Hours usually worked per week represent an extremely important variable in the calculation of pay gap statistics because this is the variable used to determine hourly wages, which is the basis for comparing the wages of females and males. In order to minimise the share of unknown values, it was decided to use the number of hours worked during the survey week in the case of persons who in the ELFS responded to the question about the number of hours usually worked per week with “hard to say, hours worked vary considerably from week to week or from month to month”. A more detailed variable analysis is available in the compatibility analysis of used data sources^a.

Table 2 shows the total number of entries and employed persons in the combined database. The table also indicates the number of people for whom it was possible to determine the number of hours usually worked per week, and the number of people for whom it was impossible.

Table 2. Presence of the variable of number of hours usually worked per week in the combined database, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Entries	29,793	31,284	29,131	27,466	27,604	28,734	31,616	31,971
Employed persons	16,803	18,092	16,933	14,340	13,690	15,161	17,443	17,915
Number of hours usually worked per week present	16,799	18,087	16,919	14,334	13,665	15,030	17,250	17,727
Number of hours usually worked per week not present	4	5	14	6	25	131	193	188

Problems occur with the variable of company size. The number of employees in the company is collected differently in the ELFS and the ESS. Both surveys do include the question “How many employees are there in this company/body/unit?”, but they have different reply options. At the same time, the answers of both surveys could be re-coded, so that there would be a single variable in the combined database. Table 3 shows how the variable of the number of employees in the ELFS and the ESS has been re-coded for the combined database.

^a <http://www.stat.ee/dokumendid/78174> (only in Estonian)

Table 3. Re-coding the variable of number of employees in the company

Combined database	ELFS	ESS
1. 1 to 10	1. 1 to 10	1. 1 to 10 5. Not certain, but fewer than 11
2. 11 to 19	2. 11 to 19	2. 11 to 19
3. 20 to 49	3. 20 to 49	3. 20 to 49
4. 50 or more	4. 50 to 99 5. 100 to 199 6. 200 to 499 7. 500 to 999 8. 1,000 or more	4. 50 or more
5. Unknown		6. Not certain, but more than 10

As one option provided in the ESS is “not certain, but more than 10”, the size of enterprise remains unknown more frequently in the ESS than in the ELFS (see Table 4).

Table 4. Presence of the variable of number of employees in the company in databases by survey, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
ELFS								
Total employed persons	9,834	11,547	11,058	8,791	8,562	9,819	11,584	11,628
Number of employees in the company unknown	454	545	94	94	113	106	144	147
ESS								
Total employed persons	6,969	6,545	5,875	5,549	5,128	5,342	5,859	6,287
Number of employees in the company unknown	423	448	283	238	198	258	258	234

The two surveys also have different methods for collecting data on the employees’ monthly wages. The ESS asks the respondent about their average net monthly wages at a particular job. If the respondent cannot or does not want to answer that question, the interviewer can mark “Does not know” or “Refuses to answer”.

In the ELFS, at first the respondent is asked about their monthly net or gross wages at their main job in the previous month. Then the respondent is asked whether the previous earnings were the same as the usual monthly wages at that job. If the previous wages were different from the average, the interviewer will ask about the respondent’s usual net or gross wages at their main job. The interviewer may note “Does not know” or “Refuses to answer” for both wage-related questions if the respondent does not want to or cannot answer the question. If the respondent does not want to respond to the question about the previous wages, the interviewer will ask about the range of net wages for the previous month.

As the ESS does not include a question about the wages received in the previous month, the variable of usual monthly wages is used in the combined database. Thus, the variables collected in the ELFS are re-coded as follows:

- if the respondent said that the previous month’s wages differed from the usual wages, the question about the usual wages is used;
- if the respondent said that the previous month’s wages were the same as usual, the net or gross monthly wages of the previous month at main job are used;
- if the respondent said their gross wages, but not net wages, the gross wages are converted to net wages, taking only income tax into account.

Ultimately, the field of average net monthly wages remains empty for persons who do not wish to or do not know how to answer the wage-related question. In their case, net wages are calculated by statistical methods through other variables, i.e. net wages are imputed. The methodology of imputation is described in chapters “Imputing wages in the ELFS” and “Imputing wages in the ESS”.

Variables collected from dissimilar groups

Some of the variables in the ELFS and the ESS are collected from dissimilar groups. Namely, in the ELFS, questions about the presence of managerial responsibilities, company sector and form of ownership are asked from all people who work, but in the ESS only from wage employees. Thus, in terms of these variables, only the responses of wage employees are taken into account. Also, the ELFS asks the size of the company from all persons who work, but the ESS of 2011 and previous surveys did not ask the size of the company from self-employed persons and unpaid family/farm workers. Thus, with regard to the number of employees shown in the combined database, it was decided to attribute the value “1” to self-employed persons and “Unknown” to unpaid family workers in the case of Estonian Social Surveys of 2011 and earlier.

Variables in the combined database

The final combined database file contains a total of 61 variables that rely on data collected by the ELFS and the ESS. An exhaustive list of the variables collected by individual surveys is provided in Annex 1 at the end of this report.

Unknown values

In the case of data collected via individual surveys, it is not always possible to gather all the information from the respondents and therefore some of the variables remain unknown (if the respondent does not want to or cannot answer). If the final response is either “Do not know” or “Refuse to answer”, then these will be treated in further analysis as unknown values.

Background variables, such as the respondent’s sex, age, ethnic nationality, marital status, place of residence, etc. should be marked for all respondents, but for many other variables values might also be unknown. In the combined database, the number of unknown values is even higher for some variables because of data processing: in order to treat the variables of two surveys as equivalent, it is sometimes necessary to re-code the variables and thus it may happen that some values have to be re-coded as unknown values. A detailed analysis of unknown values in the combined database is available in the compatibility analysis of used data sources^a.

Exceptional cases in the combined database

As the questionnaires and manner of presenting questions in the ELFS and the ESS are not always identical, there were exceptional cases that had to be taken into account when creating the databases of the two surveys.

The first central issue concerns the age groups of combined databases: in the ELFS, the respondents of the individual questionnaire included all people aged 15 to 74. In the ESS, people at the age of 15 and older responded to the individual questionnaire, but the databases that were also used when creating the combined database are adapted to the needs of Eurostat, i.e. they include people aged 16 and older.

Another pervasive nuance that needs to be taken into account in the case of ESS data is the imputed income variables. Since the ESS of 2007, the income variable has been imputed for all respondents who should have responded to the individual questionnaire, but failed to do so (refusal or loss). The imputation of these income variables in the ESS is necessary in order to calculate the income of the entire household. Considering that upon creation of the combined database, the main emphasis with regard to data retrieved from the surveys is placed on background variables which are primarily

^a <http://www.stat.ee/dokumendid/78174> (only in Estonian)

related to education and employment, these so-called “replacement questionnaires” have been omitted; neither are they included in calculating the weights of the combined database.

Table 5. Final number of entries in the combined database by survey, 2006–2013

Year	Social Survey	Labour Force Survey
2006	13,007	16,786
2007	11,875	19,409
2008	10,761	18,370
2009	11,220	16,246
2010	11,114	16,490
2011	11,076	17,658
2012	11,775	19,841
2013	12,348	19,623

The exceptional cases that occurred during the creation of the combined database:

- Although the surveys ask for the respondent’s personal identification code, there are still forms where this field has been left empty and it has not been possible to add it retrospectively during processing. In order to specify personal identification codes, it is necessary to know the person’s name and date of birth, thus it is impossible without having exact data. The greatest problems with regard to lack of personal identification code occurred in the database of the ELFS of 2006, where personal identification code was missing for 4,426 persons (see Table 6).

Table 6. Personal identification code by survey, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Code present	25,345	30,835	28,389	27,152	27,275	28,448	31,423	31,821
Code missing in the ELFS	4,426	373	701	252	288	262	168	129
Code missing in the ESS	22	76	41	62	38	24	25	21

- The wages question which forms the basis of the combined database was not included in ESS of 2006, leading to a lack of wages data for that year. The question about net monthly wages (“What are your average net monthly wages on this job?”) has been present in the individual questionnaire since 2007.
- In 2006–2008, there were problems with the place of residence variable collected in the ELFS, because place of residence is registered in the person-based database only on the county level, including the city of Tallinn, which is included under Harju county. Such a presentation of data did not allow referring to Tallinn separately or finding the place of residence of a household based on population density, which requires a three-digit rural municipality code. In order to solve that problem, the place of residence of all individual questionnaire respondents was retrieved from the original file of the relevant year and Tallinn was attributed a separate code. As the combined database requires the place of residence of household, not that of the person, the place of residence of the head of the household was taken as the place of residence of the household. If the head of household was living abroad, the place of residence of the household was determined based on other members.
- The ESS 2006–2008 did not include a question about the main language spoken at home. Therefore, these variables are present in the combined databases of these years only in the case of people who participated in the ELFS.
- ESS 2006 did not include a question about the form of ownership of the company/body, therefore it lacks the division into public and private sector. For that year there is also no data about the type of employment contract for the participants of the ESS. These variables are present in the combined database only in the case of people who participated in the ELFS.

- As the Social Surveys of 2006 and 2007 did not ask whether the respondents resided with their legal spouse, problems occurred in the combined database when determining the actual marital status of the respondent. Thus, the data of household relations matrix were used for calculating this variable.
- In 2006 and 2007, the education question was asked slightly differently in the individual surveys than in later years, which may have caused some shifts in ISCED97 levels 3b, 3a and 4b in transition to 2008 (see Table 7).

Table 7. Education level of survey participants by ISCED97 classification, 2007, 2008

(per cent)

	2007		2008	
	ELFS	ESS	ELFS	ESS
Level 0	0.5	0.5	0.4	0.4
Level 1	4.8	5.0	4.4	4.8
Level 2c	0.1	0.5	0.1	0.3
Level 2a	20.9	22.4	21.1	22.4
Level 3c	4.9	5.7	4.1	5.0
Level 3b	7.7	8.3	3.1	4.9
Level 3a	32.6	31.2	41.4	38.3
Level 4b	5.0	6.5	1.0	2.6
Level 5b	9.6	7.8	8.7	8.1
Level 5a	13.8	11.7	15.6	12.9
Level 6	0.2	0.3	0.1	0.3
Unknown	0.0	0.0	0.0	0.0

- As for the question about the number of employees in the company, the options in the ELFS 2006 and 2007 included “Does not know” and thus, the number of unknown values is significantly higher in these years than later on (value unknown for 454 persons in 2006, 545 in 2007, but only 94 in 2008).
- In the ELFS of 2006 and 2007, the marital status at the time of survey was left unmarked in the case of respondents younger than 16 years of age because it was not asked from them in the questionnaire. In the combined database, such persons are coded under the category “Single”.
- Unlike the subsequent surveys, the Social Surveys of 2006 and 2007 did not include a separate question about managerial responsibilities. However, the individual questionnaire for these years contains a question concerning the number of subordinates of the respondent, and thus this variable is used for the databases for these years. If the respondent had at least one direct subordinate, the respondent was considered to have managerial responsibility.

In 2006–2008, the format of the ELFS databases differed from the format of later ones. Namely, while the survey data of one year are usually stored in three separate files (on household, household member and individual level), the data for these years are stored in a single file, with the data of other household members indicated behind the head of household. In order to link a household member to a personal file, it is necessary to use an identifying variable, i.e. a unique identification code. Also, one has to keep in mind that in the ELFS, the same household is interviewed in two consecutive quarters, i.e. one household participates in the survey four times over the period of a year and a half.

Since 2009, the databases of the ELFS and the ESS are more compatible because of the harmonisation of background variables in individual surveys. Further information about the harmonisation of background variables can be found in “Tausttunnuste käsiraamat” (“Manual for background variables”)^a.

In 2010, Statistics Estonia adopted a new classification of occupations ISCO-08 (International Standard Classification of Occupations). Considering that the new and the old ISCO classification are

^a <http://www.stat.ee/dokumendid/31703> (only in Estonian)

not fully compatible, the occupations stated in 2006–2009 are re-coded to suit the currently used new classification. Both the new and the previous version of the classification of occupations are available on the homepage of Statistics Estonia^a, along with a description of the transition from one classification version to another^b. As the data of the ESS had not been retrospectively re-coded, they were re-coded for the period of 2006–2009.

Quality of wage variable in individual surveys

The quality of the wages variable of the ELFS and the ESS has been analysed in depth in the publication "Palgaerinevuste statistika parem kättesaadavus: kasutatavad andmeallikad" ("Increased availability of pay gap statistics: data sources used")^c.

Adding Tax and Customs Board data to the combined database

Structure of Tax and Customs Board database

The ETCB database relies on the data collected by means of Annex 1 of the TSD-form. TSD-form is the declaration of income and social tax, mandatory funded pension and unemployment insurance premium, submitted to the ETCB by the person making the payment by the 10th day of the calendar month following the month when the payment was made. The declaration form with the annexes and the procedure for filling out the declaration is established by a Regulation of the Minister of Finance.

The ETCB database includes the following variables:

- personal identification code of payment recipient;
- registry code of the person making the payment;
- year of payment;
- month of payment;
- amount of payment by category.

In the course of anonymisation, Statistics Estonia uses the personal identification code of the payment recipient to establish the recipient's sex, date of birth, and an anonymous personal identification code, which allows combining register-based data with other data sources.

In the ETCB database, sometimes more than one entry can correspond to each person. In the initial ETCB database, one entry corresponds to the payments made to one person from one company in one month. The ETCB database includes the following entries:

- if a person has received within the same month payments from several employers, then the payments from each company are recorded under separate entries;
- if the same employer has made different types of payments to the same person within one month, then it is recorded in the database as a single entry;
- if a person has received wages from the same employer in different months, then there is one entry per month in the database;
- if a person has received different types of payments from the same employer in different months, then there is one entry per month.

Based on the payment types represented in the data of the ETCB, the statistical indicator "earned income" has been established, taking into account the following payment types:

- Payment type No. 1: Earnings;
- Payment type No. 11: Remuneration of a board member;
- Payment type No. 15: Remuneration received based on a contract under the Law of Obligations (usually in the case of a contract for services).

^a http://metaweb.stat.ee/classificator_publish_list.htm?&siteLanguage=en

^b http://metaweb.stat.ee/classificator_correspondence_list.htm?siteLanguage=ee (only in Estonian)

^c <http://www.stat.ee/dokumendid/78174> (only in Estonian)

- In the case of earned income, negative payments are also taken into account.

Adding Tax and Customs Board data to the combined database of the ELFS and the ESS

In order to link the data of the ETCB to the ELFS and ESS combined database, the ETCB database first has to be individual-based, i.e. there will be only one entry for each person.

As the ETCB database is used to produce various variables concerning the average earned income of three months (the survey month and the months before and after), it is necessary to take into account the payments made not only in the year in question, but also in the last month of the previous year and the first month of the subsequent year. For example, the database for 2012 contains all payments made in 2012, in December 2011 and in January 2013. As the ETCB database contains all persons who have received payments according to the ETCB, only the persons who participated in the survey of the respective year (either in the ELFS or the ESS) will be included. Then three types of earned income will be determined for each entry and other income will be calculated. This results in the ETCB database for the relevant year.

In order to make the ETCB database individual-based, the entries of the database have to be aggregated (at first, aggregated on the level of person, month and year). This means that there will be one entry per month for each person. If a person received payments from several employers in a month, the types of earned income per month will be added together. Then the database will be restructured so that there will be one line for each person and the payments for each month are shown as separate variables.

The ETCB database wage data can be added for all those who participated in the ELFS or the ESS and whose personal identification code is known. Table 8 shows the number of people with unknown personal identification code during the period 2006–2013, i.e. the persons for whom it was not possible to link the ETCB database wage data.

Table 8. Persons with unknown personal identification code in the combined database

	2006	2007	2008	2009	2010	2011	2012	2013
Code unknown	4,448	449	742	314	329	286	193	150

Table 9 shows the number of wage employees according to the data of the ELFS and the ESS, and the number of people whose earned income was above zero according to the data of the ETCB.

Table 9. Wage employees and earned income by survey, 2006–2013

Year	ELFS	ESS	Total
2006			
Wage employees	9,005	6,419	15,424
Earned income present	6,245	6,109	12,354
2007			
Wage employees	10,513	6,033	16,546
Earned income present	9,792	5,679	15,471
2008			
Wage employees	10,165	5,453	15,618
Earned income present	9,243	5,142	14,385
2009			
Wage employees	8,033	5,104	13,137
Earned income present	7,482	4,819	12,301
2010			

Wage employees	7,828	4,708	12,536
Earned income present	7,294	4,443	11,737
2011			
Wage employees	8,970	4,887	13,857
Earned income present	8,365	4,577	12,942
2012			
Wage employees	10,542	5,363	15,905
Earned income present	9,893	5,054	14,947
2013			
Wage employees	10,605	5,753	16,358
Earned income present	9,975	5,418	15,393

Variables produced based on the Estonian Tax and Customs Board database

Earned income

After making the database individual-based, the next step would be to calculate the necessary variables relating to different types of income, earned income and wages payments from the main employer. The variables included in the combined database are presented at the end of the report (see Annex 1).

At first, earned income variables are calculated in total and also separately by type (S_1 or earnings, S_11 or board member remuneration, S_15 or pay received based on a contract under the Law of Obligations, and MT or other income). In each year's database, earned income is calculated for these three TSD types (December of the previous year, the year in question, and January of the following year). Earned income per month is determined by adding these three payment types.

The sum of variables S_1, S_11 and S_15 is the total earned income, which is also calculated for December of the previous year (variable TT_12_EA), every month of the current year (variables TT_01 ... TT_12) and January of the subsequent year (variable TT_01_JA). Additionally, the annual total sum of all income types and earned income is calculated; this is the total of monthly earned income.

The number of months when a person received earned income is calculated by adding together the number of months when the person received earned income (i.e. the person has received at least one out of the three payment types). If the person has received negative payments in some of the months, it will also be taken into account when calculating the number of months. Also, the number of months when the corresponding payment type was received will be determined for the three payment types and other income.

Earned income paid by main employer

All variables calculated for earned income are also calculated with regard to the main employer, i.e. the employer who made the biggest payment.

The main employer is determined as follows:

- First, the main employer per month is determined: if a person received payments from several employers in one month, the main employer is the one who made the biggest payment of earned income, i.e. whose total sum of variables S_1 + S_11 + S_15 is largest.
- The main employer per year is determined based on the monthly main employers and the calendar year only (i.e. 12 months) is taken into account. The main employer is the employer who has made payments over a longer period (a greater number of months). If several employers have made payments during periods of the same length, the employer who made the greatest earned income payment (largest sum of S_1 + S_11 + S_15) will be the main employer.

After identifying the main employer, the payments made by the main employer will be found and such variables will be attributed the prefix PT (indicating the main employer). Each year's database contains separate entries for earned income categories (PT_S1, PT_S11, PT_S15) and total other income

(PT_MT), paid by the main employer in December of the previous year, each month of the current year and January of the subsequent year. It also features the calculation of earned income paid by the main employer per each month of the current year (variables PT_01 ... PT_12) and December of the previous year (variable PT_12_EA) and January of the subsequent year (variable PT_01_JA).

Besides monthly payments, the total annual earned income paid by the main employer is calculated (PT_aasta) and the number of months when it was received (PT_kuude_arv). Also, for each payment category, the annual sum and the number of months when it was paid have been calculated. The list of the database variables created based on ETCB data is in Annex 1.

Additional variables calculated based on earned income

In addition to the variables of earned income payments and payments of main employer described above, the database also includes various additional variables (see Annex 1). Based on the variables described above, the average monthly earned income per calendar year and the average hourly earned income are calculated.

In order to determine the annual average earned income, the earned income of a year is divided by the number of months when earned income was received. The average annual earned income takes into account only the persons whose average annual earned income was above zero.

Expressed as formula:

$$\text{Annual average earned income per month} = \frac{\text{Total annual earned income}}{\text{No. of months when earned income was received}}$$

Expressed using names of variables in database:

$$TT_ak = \frac{TT_aasta}{TT_kuude_arv}$$

As ELFS and ESS interview people in different months, the annual average earned income per month may not match the wages of the job the person had at the time of the survey. For example, a person might have had a low-paying job at the time of the survey, and a higher-paying job for the rest of the year. Then the annual average earned income of that person may significantly differ from the income indicated in the survey. To clarify this, the combined database has determined the following earned income variables:

- Earned income of survey month (variable TT_uk) is the earned income of the month corresponding to the survey month. E.g. if a person was surveyed in January, the total sum of payments made in January will be taken into account.
- Earned income of the month following the survey month (variable TT_ujk) is the sum of earned income of the month following the survey month. E.g. if a person was surveyed in January, this variable consists in total payments made to that person in February.
- Calculation of average earned income of three months (survey month, the month preceding it and the one following it) (variable TT_3k) consists in two phases. At first, the total earned income of three months is found for each survey month (the month preceding the survey month, the survey month, and the month following the survey month). For example, in the case of January, it will be the sum of earned income payments of December of the previous year, and January and February of the present year. Then the number of months the person received earned income out of these three months is determined. If the person received negative payments in the relevant month, these will also be included in the earned income. To find the average earned income of three months, the earned income of three months is divided by the number of months when earned income was received.

As a result of the analysis, it was decided to use the annual average earned income per month as earned income (TT_ak). A more detailed analysis is available in the compatibility analysis of used data sources^a (pp. 55–66).

^a <http://www.stat.ee/dokumendid/78174> (only in Estonian)

Besides the annual and monthly average earned income, it is possible to calculate earned income per hour. Both the number of hours indicated in the individual questionnaire (usual number of hours worked per week) and the number of prescribed hours can be used to do this. In order to calculate hourly wages, it is necessary to know the number of working days per year. The number of working days per year is determined as 253 days, which is the average of 2006–2014. An average is used in order to avoid changes in wages in case the number of working days differs year by year.

Table 10. Number of working days per year, 2006–2014

	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
Number of working days	254	254	253	252	254	255	254	253	251	253.333

Hourly wages are calculated with the following formula:

$$\text{Hourly wages} = \frac{\text{Annual average earned income per month} * 12}{(\text{Hours usually worked per week} / 5) * 253}$$

Expressed using names of variables in database:

$$TT_tund = \frac{TT_ak * 12}{(T_tund / 5) * 253}$$

In order to calculate hourly wages based on prescribed hours, the hours usually worked per week in the formula are replaced with the number of prescribed hours. Prescribed hours are calculated as follows:

- If a person works part time and the number of hours usually worked per week is less than 35, the prescribed working hours equal the hours usually worked per week;
- If a person works part time and the number of hours usually worked per week is more than 35, the prescribed working hours equal 20 hours;
- If a person works full time and is an educational worker, the prescribed working hours equal 35 hours;
- If a person works full time and is not an educational worker, the prescribed working hours equal 40 hours.

Calculation of expansion factors (weights)

To be able to expand the results of the combined database of the ELFS and the ESS to total population, it is necessary to calculate new expansion factors or weights. At the same time, this eliminates the problem of double recording of recurring persons in the ELFS, because it allows recording recurring people as participating in the survey only once.

Household weights

For the combined database, the weights are calculated based on survey-specific weights. The weight used is called WGT_Y in the case of the ELFS and DB090 in the case of the ESS. When compiling the datasets, the ESS weight is multiplied by the coefficient λ and the ELFS weight by the coefficient $1-\lambda$. Value of the coefficient λ depends on the ratio of efficient sample sizes of the surveys:

$$\alpha = \frac{n_{ESS}^{eff}}{n_{ESS}^{eff} + n_{ELFS}^{eff}},$$

$$n_{ESS}^{eff} = \frac{n}{1 + \sqrt{\text{var}(DB090) / \text{mean}(DB090)}},$$

where n is the number of households in the ESS

$$n_{ELFS}^{eff} = \frac{n}{1 + \sqrt{\text{var}(WGT_Y)/\text{mean}(WGT_Y)}}$$

where n is the number of households in the ELFS

The survey with a greater sample size (and thus smaller variation in weights) will be attributed a greater weight. For example, in 2011, $\lambda = 0.358$ and in 2012, $\lambda = 0.346$. As a result, the weights of the households taken from different surveys are similar in the common dataset.

The final weights are calibrated by 10-year sex-age groups, and county and category of settlement unit. The adjusted population figure, of which the number of persons residing in institutions has been deducted, is used.

Personal weights

Personal weights, upon the calculation of which the lack of household-based response is taken into account, are used if the analysis takes into account only the respondents who filled out the individual questionnaire. These weights are calculated based on household weights so that the total weights in the sex-age groups of one year are equal among the respondents and all persons aged 15 or older. Thus, personal weights are no longer equal within the household.

As the variables used in the combined database are collected with the individual questionnaire, only personal weight is added to the combined database.

In the established combined database there are in total two weight variables: survey weight and weight. The survey weight variable that has been added to the database allows observing survey data separately by survey. The weight variable, on the other hand, covers the combined database, i.e. it allows conducting analyses based on the entire combined database.

Imputation of unknown values

Some variables, such as wages, number of hours worked and working time format, are extremely important in the calculation of the pay gap. In order to reduce the share of unknown values for these questions, such unknown values are provided with a potential estimate, i.e. unknown values are imputed.

Imputation of wages in the ELFS

In the ELFS, all employees are asked about their exact monthly wages for the previous month. To avoid confusion with regard to distinguishing between gross and/or net wages, the respondent is asked to indicate both. Those who cannot or do not want to say their exact wages are asked to indicate the approximate range of their net wages for the previous month. For those respondents who choose not to disclose their wages in the survey, the wages will be determined on the basis of other variables, i.e. for those people, net monthly wages are imputed.

At first, the net monthly wage range is imputed for all those employees who did not want or could not disclose it. The range is imputed by applying randomized logistic regression in IVEWARE software. Variables used in the model include the person's sex, age group (5-year age groups), county of residence and category of settlement unit of the place of residence.

Then, the exact net monthly wages are imputed for all employees according to their wages range by using the hot-deck method, where a random person is selected from the same wages range and his or her exact wages value is copied.

Considering that imputation uses net wages, net wages are present for all employees in the database. Gross wages used in the combined database are calculated using the formula below:

If net wages \leq 144, then gross wages = net wages

If net wages $>$ 144, then gross wages = (net wages – 0.21 * 144) / (1 – 0.21) = (net wages – 30.24) / 0.79

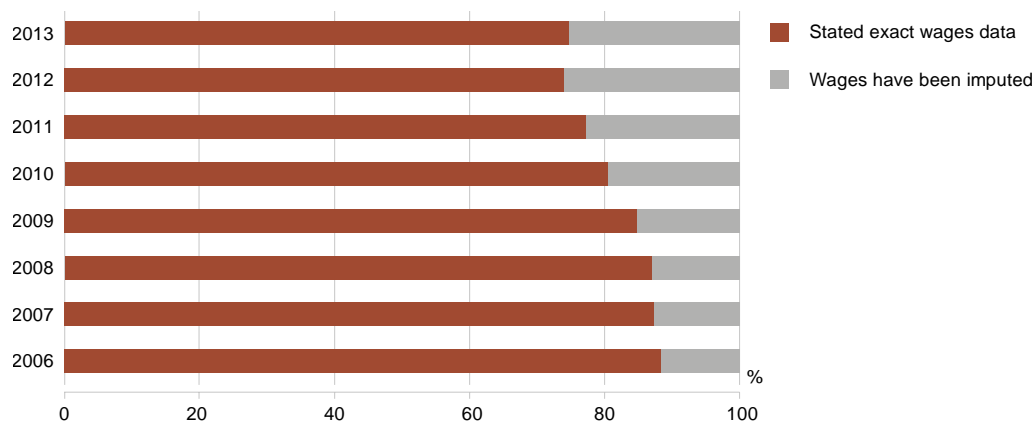
A more detailed explanation concerning the conversion of net wages to gross wages by year is available in Table 11. The conversions only take income tax into account. Amounts are given in kroons for 2006–2010 and in euros for 2011–2013.

Table 11. Converting net wages to gross wages, 2006–2013

Year	Non-taxable, kroons/euros	Tax rate, %	Tax rate* non-taxable	1 – tax rate	Formula
2006	2,000	24	480.0	0.76	$B = (N - 480) / 0.76$
2007	2,000	22	440.0	0.78	$B = (N - 440) / 0.78$
2008	2,250	21	472.5	0.79	$B = (N - 472.5) / 0.79$
2009	2,250	21	472.5	0.79	$B = (N - 472.5) / 0.79$
2010	2,250	21	472.5	0.79	$B = (N - 472.5) / 0.79$
2011	144	21	30.2	0.79	$B = (N - 30.24) / 0.79$
2012	144	21	30.2	0.79	$B = (N - 30.24) / 0.79$
2013	144	21	30.2	0.79	$B = (N - 30.24) / 0.79$

Figure 2 indicates the share of surveyed employees who provided their exact wage data and the share of employees whose wages have been imputed.

Figure 2. Wages data of the Estonian Labour Force Survey, 2006–2013



Imputation of wages in the ESS

The Estonian Social Survey also asks from all employees about the earnings they received for the previous month. As for the wages received at the main job, only the net amount is asked while gross value is calculated later in the course of processing. Wage variables are imputed for those persons who refused to disclose their wages or do not know their wages.

Ultimately, the question regarding average net wages was left unanswered in the case of those persons who did not want or could not answer the wage question. In their case net wages are determined by statistical methods through other variables, i.e. their net wages are imputed.

Wages variable is imputed by randomized linear regression with the IweWare software and the imputation model used contains the following variables:

- average monthly wages of the previous year and quartile thereof;
- person's age, age squared and sex;
- indicator showing whether the person worked at least 11 months in the previous year;
- indicator showing whether the person has managerial responsibilities;
- indicator showing whether the person changed jobs within the previous 12 months;
- indicator showing whether in the unit where a person works there are more than 10 employees;
- indicator showing whether the person's profession is in the field of service and sales.

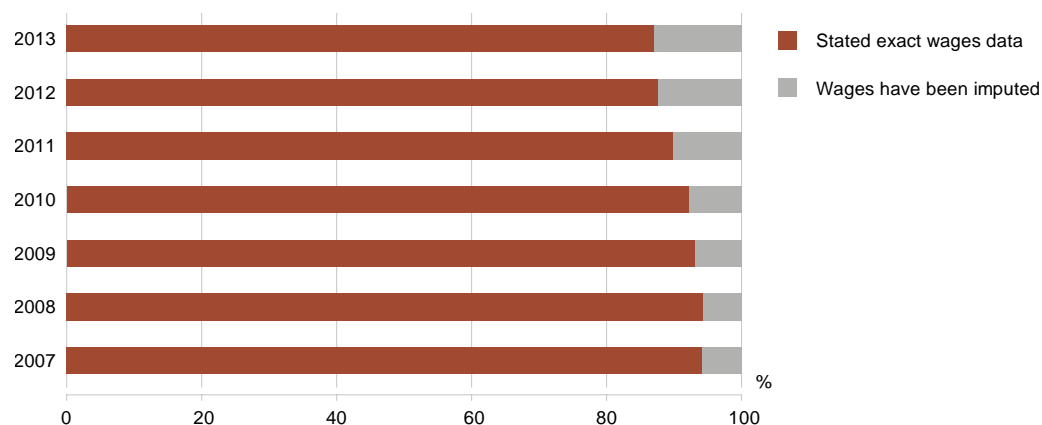
As it was decided to add a gross wage variable to the combined database, imputed net wages are then converted to gross wages as follows:

If net wages ≤ 144 , then gross wages = net wages

If net wages > 144 , then gross wages = $(\text{net wages} - 0.21 * 144) / (1 - 0.21) = (\text{net wages} - 30.24) / 0.79$

A more detailed explanation about the conversion of net wages to gross wages by year is presented in Table 11.

Figure 3. Wage data of the Estonian Social Survey, 2007–2013



Imputation of number of hours worked

The analysis of output indicators revealed that the database contained illogical values regarding the number of hours worked. Therefore, it was decided to impute new values to replace the values that seemed illogical. For that purpose, the Estonian average hourly gross wages were calculated by economic activity, and those wages were then compared to the hourly wages calculated based on the number of hours stated in the survey.

To start with, Estonian average hourly gross wages by main economic activity (EMTAK 2008) of both full-time and part-time employees for 2008–2013^a were found based on data from the database of Statistics Estonia. Then different shares – 10%, 20% and 30% – of average hourly gross wages of full-time and part-time employees were calculated. These shares were used to calculate, by economic activity, average hourly gross wages ranges within which hourly gross wages were allowed to vary. After analysing the results it was decided that hourly wages could differ from the average hourly gross wages by approximately 30%. These limits were used to find whether the number of hours worked needed to be imputed or not.

Considering that in 2006 and 2007, the version of EMTAK classification was not yet that of 2008 but rather an older version^b was being used, the hourly wages for those two years had to be calculated separately. As the average hourly gross wages by economic activity were known for 2006 and 2007, a coefficient, indicating the ratio of the difference between the average hourly gross wages of a given economic activity of a year and those of the following year to the hourly gross wages of the following year, was determined first. Then the hourly gross wages were multiplied by that coefficient and then the result was deducted from the hourly gross wages of the relevant economic activity of the subsequent year. For example, full-time employees' average hourly gross wages were 5.02 euros in 2008 and 4.36 euros in 2007. The coefficient value equals $(5.02 - 4.36) / 5.02 = 0.13$. In 2008, the average hourly gross wages of full-time employees working in agriculture, forestry and fishing were 4.06 euros. Thus, the average hourly gross wages for the same economic activity in 2007 were calculated as follows: $4.06 - (4.06 * 0.13) = 3.53$ €. In 2006 as well the coefficient was calculated first, followed by the calculation of average hourly gross wages of economic activities in comparison with 2007. Annex 2 indicates the 30% ranges of hourly gross wages of full-time and part-time employees by economic activity.

Persons whose number of hours worked will be imputed are found in two ways. One group of respondents said that it was hard to say the number of hours usually worked per week because the number varied considerably from week to week or from month to month. The number of hours worked was imputed for those respondents. The second group was found by using the 30% ranges of hourly

^a Statistics Estonia database, table WS5212: Average hourly gross wages (salaries) by economic activity (EMTAK 2008)

^b Statistics Estonia database, table WS5312: Average hourly gross wages (salaries) by economic activity (EMTAK 2003) (2000–2008)

gross wages calculated above. Based on that, the number of hours worked were imputed if the person's hourly wages (according to ETCB data) did not fall in the 30% range of average hourly gross wages (i.e., the hourly gross wages were above or below that range). The number of hours worked was only taken into account if the number was either up to 10 or at least 60. Thus, the number of hours worked was imputed if the following criteria were met:

- the person was employed (i.e. I_TOOTAV = 1);
- the person's number of hours worked ranged from 0 to 10 (i.e. T_TUND > 0 & T_TUND ≤ 10) or was at least 60 (i.e. T_TUND ≥ 60);
- the person's hourly gross wages did not fall in the 30% range of Estonian average hourly gross wages.

As a result, there were persons whose number of hours worked had to be imputed again. Taking into account such persons and the persons who refused to provide the number of hours worked, Table 12 presents the total number of persons whose number of hours worked had to be imputed again.

Table 12. Imputing the number of hours worked, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Persons whose number of hours was imputed	283	268	260	245	298	383	517	494

At first, the ranges of hours worked are imputed by means of randomized logistic regression with the software IweWare and then exact values within each range are imputed by using the hot-deck method. Upon the imputation of ranges the following auxiliary variables are used: indicators of groups of occupation and economic activities, presence of off-hour job, age group, total annual earned income, employment status, wage employee/not wage employee, presence of managerial responsibility.

Combined database variables indicate the initial (survey-based) number of hours usually worked per week (T_TUND), the number of hours worked with imputed values (T_TUND_imp) and a variable that indicates whether the number of hours worked was imputed or not (t_tund_i).

Imputation of working time format

As was the case with the number of hours worked, the database contained illogical results with regard to the working time format (full-time or part-time employment) as well.

In order to know the number of people whose working time format must be imputed, it is necessary to determine the total number of persons, in whose case the working time format is unknown already in the ELFS or the ESS database, and add that number to the number of persons, whose working time format is illogical. Thus, working time format was imputed if the following conditions were met:

- the person was employed (i.e. I_TOOTAV = 1);
- the person worked full time and the number of hours worked was in the range 0–10 (i.e. T_AEG = 1 & (T_TUND > 0 & T_TUND ≤ 10)) or the person worked part-time and the number of hours worked was at least 60 (i.e. T_AEG = 2 & T_TUND ≥ 60).

Table 13 presents the number of persons for whom the working time format was imputed in 2006–2013.

Table 13. Imputation of working time format, 2006–2013

	2006	2007	2008	2009	2010	2011	2012	2013
Persons whose working time format was imputed	0	2	6	7	7	14	15	20

Working time format is imputed by using randomized logistic regression and IVEWARE software. Imputation of variable T_AEG involves the following auxiliary variables: indicators of groups of occupation and economic activities, presence of off-hour job, age group, total annual earned income.

The combined database variables present the initial (survey-based) working time format (T_AEG), the working time format with imputed values (T_AEG_imp) and a variable indicating whether the working time format is imputed or not (t_aeg_i).

Summary of establishing the combined database

Within the framework of the project “Increased availability of gender pay gap statistics”, combined databases were created for the years 2006–2013 (8 databases in total). The databases of two individual surveys – the Estonian Labour Force Survey and the Estonian Social Survey – were combined. They were supplemented with data from database of the Estonian Tax and Customs Board. In order to be able to expand the data to total population, new expansion factors or weights were determined for the combined databases.

The final combined database contains a total of 246 variables: 61 variables based on the ELFS and the ESS, 160 variables based on the data of the Estonian Tax and Customs Board and 25 additional variables determined on the basis of data from both the individual surveys and the Tax and Customs Board data (see Annex 1).

PAY GAP AND INDICATORS THAT AFFECT IT

In the framework of the project “Increased availability of gender pay gap statistics”, gender pay gap is published by background variable along with several additional indicators. These indicators are based on the recommended indicators established by the political studies centre PRAXIS and the Centre for Applied Research CentAR. In order to understand the dynamics of the gender pay gap, it is necessary to simultaneously monitor the development of society and labour market with regard to the aspects behind the occurrence and persistence of the gender pay gap^a. The final list of indicators and methods for their calculation was agreed on with the experts during a workshop held in Statistics Estonia in March 2015. The workshop materials are available on the webpage of Statistics Estonia^b.

Various indicators along with a short description are presented below.

Average gender pay gap

Shows the ratio of the difference between the hourly wages of males and females to the hourly wages of males.

$$\frac{\text{hourly gross wages of males} - \text{hourly gross wages of females}}{\text{hourly gross wages of males}} \times 100\%$$

Data source: combined database

Period: 2006–2013

Unit: %

The gender pay gap is calculated based on the following background variables:

- Age
- Ethnic nationality
- Place of residence
- Level of education
- Field of highest level of education attained
- Occupation
- Economic activity
- Working time format
- Presence of managerial tasks
- Type of employer
- Size of enterprise/body
- Actual marital status
- Presence of preschool age children

Share of part-time employees among female and male wage employees

Shows how big is the share of males/females that work part-time or full-time.

Data source: Estonian Labour Force Survey

Period: 2006–2013

Unit: %

^a https://www.sm.ee/sites/default/files/content-editors/Ministeerium_kontaktid/Valjaanded/toimetised_20112.pdf (only in Estonian)

^b <http://www.stat.ee/183300> (only in Estonian)

Sex segregation on the labour market

Shows the share of females that need to change jobs in order to ensure that the shares of males and females would be equal in all occupations and economic activities. The Duncan and Duncan index is used for calculation. The segregation index can be calculated based on both occupation and economic activity.

Calculation formula:

$$S_t = \frac{1}{2} \sum_i \left(\frac{q_{it} T_{it}}{\sum_i q_{it} T_{it}} - \frac{p_{it} T_{it}}{\sum_i p_{it} T_{it}} \right)$$

Where $p_{it} = F_{it} / T_{it}$ and $q_{it} = (1 - p_{it}) = M_{it} / T_{it}$

F_{it} = number of females in economic activity i and in year t

M_{it} = number of males in economic activity i and in year t

$T_{it} = F_{it} + M_{it}$ = total number of employees in economic activity i and in year t

Shortened version:

$$S_t = \frac{1}{2} \sum_i \left(\frac{M_i}{M} - \frac{F_i}{F} \right)$$

Data source: combined database

Period: 2006–2013

Unit: %

Share of men and women among managers

Shows how big is the share of men and how big is that of women among managers. Managers are persons who are classified as managers according to ISCO and who have managerial responsibility.

Data source: combined database

Period: 2006–2013

Unit: %

Share of men and women among entrepreneurs

Shows how big is the share of men and how big is that of women among entrepreneurs. Entrepreneurs are those employed persons who identify themselves as self-employed persons, or employers or farmers with paid employees.

Data source: combined database

Period: 2006–2013

Unit: %

Passivity caused by unavailability of child care services

Shows the share of people who work part time or are not seeking a job due to lack of child care services. Respondents who work part time or do not seek a job are asked about the main reason for working part time or not seeking a job. If the main reason is parental leave or the need to take care of children or adults who need taking care of then the respondents are asked whether they are working part time or not seeking a job because the suitable child care service is unavailable or too expensive.

Data source: the Estonian Labour Force Survey

Period: 2006–2013

Unit: %

Share of men among recipients of parental benefit

Shows how big is the share of men among the recipients of parental benefit. It is also possible to calculate the average parental benefit by sex.

Data source: Tax and Customs Board, Register of taxable persons

Period: 2006–2013

Unit: %, euros

Share of men among users of parental leave

Shows how big is the share of men among the users of parental leave.

Data source: State pension insurance register of the National Social Insurance Board

Period: 2006–2013

Unit: %

Share of men who have used paternity leave in the number of children born

Shows the share of men who have used paternity leave in all live births in one year. This is calculated by dividing the number of men who used paternity leave in one year by the number of children born that year and then multiplying the result by 100.

Data source: State pension insurance register of the National Social Insurance Board

Period: 2006–2013

Unit: %

Time use of men and women

Time use shows the average length of paid and unpaid work by sex i.e. the number of minutes men and women spend on unpaid and paid labour per day.

Data source: Time Use Survey

Period: 2000, 2010

Unit: minutes

Impact of parenthood on the employment gap

Shows the employment gap of men and women by presence of underage children. Employment gap is the difference between the employment rates of men and women in percentage points.

Data source: Estonian Labour Force Survey

Period: 2006–2013

Unit: percentage points

Share of companies founded by men and women among new companies and share of men and women among sole proprietorships

Shows the share of companies founded by men and women among new companies. The share of men and women among registered sole proprietorships will also be calculated.

Data source: Commercial register

Period: 2006–2013

Unit: %

COMPARISON BETWEEN THE DATA OF THE COMBINED DATABASE AND THE DATA OF THE STRUCTURE OF EARNINGS SURVEY

At the moment, Statistics Estonia uses data of the Structure of Earnings Survey (SES) to calculate the gender pay gap. The survey is conducted every 4 years, the last one took place in 2014. The purpose of the survey is to provide accurate and harmonised data about the earnings of the residents of EU Member States. The survey provides information about employees' earnings, various individual characteristics (e.g. sex, age, occupation, length of service, the highest level of education attained, etc.) and employer's characteristics (e.g. level of economic activity, size and location of company)^a. In 2014, the number of objects of the Structure of Earnings Survey in the population was 80,500 and in the sample the number was 10,900. Data collection takes place via the online electronic data submission channel eSTAT. eSTAT is used for monitoring the receipt of questionnaires. Questionnaires are designed for independent completion online in a web environment, they include instructions and verifiers^b.

In the years between the surveys, Member States have to decide on their own how they produce pay gap statistics. In Estonia, this is done by using the questionnaire form No. 1410 "Wages and salaries, annex for October"^c. The number of objects (companies) in the population in 2013 was 81,361 and the number of objects in the sample was 11,592. State and local government authorities and companies with at least 50 employees are subject to exhaustive surveying, companies with fewer than 50 employees are subject to stratified simple random sampling by economic activity^d. Wage statistics population, sample, number of respondents and response rate by main economic activity (EMTAK 2008) are shown by year on the website of Statistics Estonia^e. Companies included in the sample will fill out a questionnaire about their employees based on the month of October. The questionnaire provides information about the number of employees receiving pay, the number of hours worked in October and pay received for time actually worked by sex (see Table 14).

In annex for October, companies included in the sample are asked to submit information about the wages and working hours of their employees on the company level, which means that the data of all employees of the company are aggregated. As a result, identical hourly wages will be attributed to the females and males working in this company. For example, if the company has 10 female employees who earn different monthly wages and have different workloads, their monthly wages and hours worked will be aggregated. Then the sum of monthly wages is divided by the total number of hours worked, resulting in the average hourly wages of females in this company, which is then attributed to all the females working in this company. The same methodology is used to calculate the average hourly wages for males in each particular company. As there are significantly more males in the high-earning population (according to both the ELFS and the ESS of 2012, males comprised 80% and females only 20% or one fifth in the top decile of earnings, while the situation was reversed in the first or lowest decile of earnings – for details, see the compatibility analysis of used data sources^f), such a methodology of pay gap calculation increases the average hourly wages of men.

^a http://ec.europa.eu/eurostat/web/microdata/structure_of_earnings_survey

^b <http://www.stat.ee/esms-metaandmed?id=68932&code=21102> (only in Estonian)

^c http://www.stat.ee/public/files/aruandevormid/2015/21108141015_Questionnaire_as_a_PDF_file_en.pdf?t=1442993628

^d <http://www.stat.ee/68932?id=68932&code=21108> (only in Estonian)

^e Statistics Estonia database table WS5217: Population and sample of wages (salaries) statistics by economic activity (EMTAK 2008)

^f <http://www.stat.ee/dokumendid/78174> (only in Estonian)

Table 14. Questionnaire No. 1410 “Wages and salaries”, Annex for October

WAGES AND SALARIES, ANNEX FOR OCTOBER	
1	Number of employees who were listed in the staff of an employer and who received pay in October - males
2	Number of hours worked in October – males (expressed in thousands of hours to two decimal places)
3	Pay for time actually worked - males
4	Number of employees who were listed in the staff of an employer and who received pay in October - females
5	Number of hours worked in October – females (expressed in thousands of hours to two decimal places)
6	Pay for time actually worked - females

In the EU, the gender pay gap is measured and monitored with an indicator in unadjusted form, meaning that the calculation of the wage gap does not take into account individual characteristics of the employee, which might explain the wage gap. The gender pay gap in unadjusted form indicates an overall inequality between women and men, caused by various factors:

- females and males work in different economic activities and occupations;
- the organisation of working time of females differs from that of males;
- the so-called women’s and men’s jobs are remunerated differently.

Thus, the gender pay gap is associated with many factors arising from legislation but also with social and economic factors, which is ultimately a much broader topic than just the idea of equal work and pay.

The calculation of the gender pay gap published by Eurostat takes into account all employees irrespective of their age or number of hours worked. However, it does not include all economic activities, nor companies with fewer than 10 employees. Data on economic activities are collected based on the EMTAK 2008 classification (The Estonian Classification of Economic Activities) and it does not include employees engaged in the following economic activities:

- Agriculture, forestry and fishing (economic activity alphabetic code A);
- Public administration and defence; compulsory social security (economic activity alphabetic code O);
- Activities of households as employers; undifferentiated goods and services producing activities for households for own use (economic activity alphabetic code T);
- Activities of extraterritorial organisations and bodies (economic activity alphabetic code U).

Employees’ wages are comprised of gross wages, bonuses and allowances paid regularly and remuneration for overtime. They do not include allowances paid irregularly. Working hours are based on the number of hours worked stated by the employer. This methodology remains the same both in the Structure of Earnings Survey years and in the interim period when gender pay gap is calculated based on the annex for October.

Statistics Estonia also publishes gender pay gap data (Table PA5335: Gender pay gap by economic activity (EMTAK 2008), October), which rely on the same data source as the Eurostat indicator, but its calculation methodology differs significantly. The pay gap published in the Eurostat database is bigger than the one published in the Statistics Estonia database, for example in 2013, the pay gap in Estonia was 29.9% according to Eurostat data, and 24.8% according to Statistics Estonia data. This has two main methodological reasons. Firstly, the methodology of Eurostat takes into account only companies with more than 10 employees, while Statistics Estonia includes companies with fewer than 10 employees as well. Considering that there are many small enterprises in Estonia, i.e. companies with fewer than 10 employees in which the wages of men and women might not differ substantially, the pay gap published by Statistics Estonia is smaller than that published by Eurostat. Secondly, based on the EMTAK 2008 classification of economic activities, Eurostat only takes into account economic activities

B–S, except O. This leaves out economic activities A (agriculture, forestry, and fishing) and O (public administration and defence; compulsory social security). However, Statistics Estonia takes all economic activities A–S into account. As the pay gap is generally smaller in economic activities A and O than in other economic activities, the omission of these economic activities also affects the size of the pay gap.

Eurostat publishes its gender pay gap in unadjusted form in the Eurostat database^a under Population and social conditions → Labour market (labour) → Earnings → Gender pay gap in unadjusted form – NACE Rev. 2. Statistics Estonia publishes its gender pay gap in the Statistics Estonia database under Economy → Wages and salaries and labour cost → Wages and salaries → Annual statistics in table WS5335: Gender pay gap by economic activity (EMTAK 2008), October.

Conducting the SES every four years (in 2010, 2014, etc.) is mandatory for all European countries. In the intermediate years, it is up to each country to decide which data source it will use for collecting pay gap data. Considering that the wages survey annex for October used in Estonia is very labour-intensive, the project work group started to look for possibilities to reduce the administrative burden of respondents by using already existing data sources and register data. Together with the expert group convened under the project, a decision was taken in the workshop in December 2013 to use annual individual surveys (the ELFS and the ESS) as alternative data sources, to which exact wage data from the database of the Estonian Tax and Customs Board will be linked. Calculation of the pay gap based on individual surveys also allows including various background variables (e.g. level of education, field of highest level of education attained, marital status, presence of small children, etc.) that enable in-depth analysis of the pay gap and its dynamics.

The gender pay gap calculated on the basis of the Structure of Earnings Survey and the wages survey annex for October and the one calculated on the basis of the combined database compiled under the project are not identical because they are calculated on the basis of different data sources. Another significant dissimilarity is that the SES and the wages survey annex for October are company-based, while the combined database is person-based. This means that in the SES and the wages survey, the company submits data on its employees' wages and number of hours worked for October. As in the individual surveys, the respondents are asked to indicate the number of hours actually worked, such an approach will show how many hours the respondents actually work (which is something the employer might not know exactly). Therefore, it is possible to use two approaches to calculate the pay gap: the company-based approach, where the company responds about its employees' number of hours worked, and the person-based approach, where the person gives the number of hours he or she worked per week. It is impossible to say that one approach is right and the other wrong, they are just different.

Another difference between the wages survey annex for October and the combined database is that, as individual surveys ask respondents about various background variables, it is possible to analyse the pay gap in depth and to monitor its changes in time. However, what is problematic about the combined database of individual surveys is that in some economic activities, the number of respondents is rather small, which is why it is not possible to get reliable data on these economic activities.

Figure 4 shows the gender pay gap calculated by Eurostat, Statistics Estonia and combined database of the individual surveys in the years 2006–2013. Statistics Estonia has no data for 2009 because neither the Structure of Earnings Survey nor the wages survey annex for October was conducted that year. The figure shows that the indicators of Eurostat and Statistics Estonia are moving in the same direction every year, but they are also calculated on the basis of the same data. The combined database indicators coincide more with those published by Statistics Estonia than with those of Eurostat because the first two have the same calculation methodology (it includes small enterprises and all economic activities). The greatest differences in pay gap indicators occurred in 2006 and 2007. The reason could be that in those years, very many respondents of the individual surveys failed to indicate their personal identification number and it was not possible to identify it in retrospect.

^a <http://ec.europa.eu/eurostat/data/database>

Therefore, it was impossible to add their wage data from the register of taxable persons of the Estonian Tax and Customs Board or involve them in the pay gap analysis.

Figure 4. Gender pay gap by Eurostat, Statistics Estonia and the combined database, 2006–2013

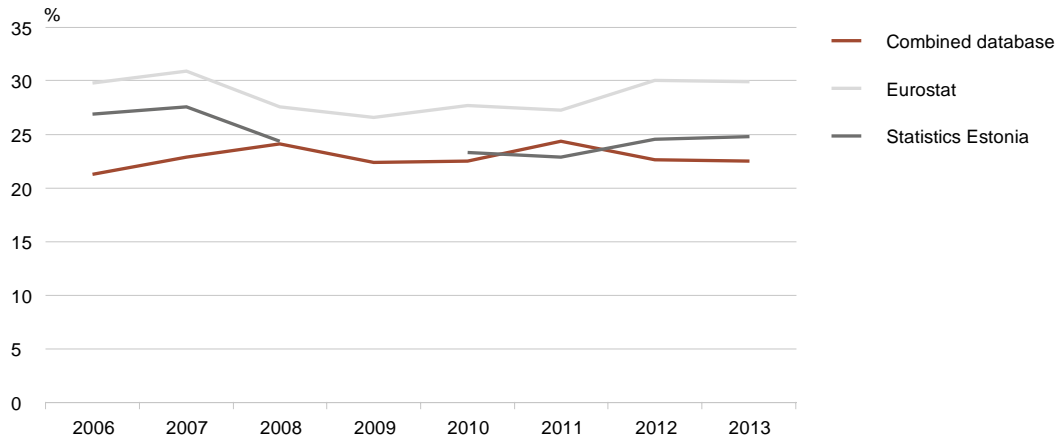


Figure 5 shows the pay gap by economic activity calculated based on the data of Eurostat, Statistics Estonia, and the combined database created under the project (see Annex 3).

Figure 5. Gender pay gap by economic activity, calculated based on the data of Eurostat, Statistics Estonia, and the combined database, 2013



STRATEGY OF PAY GAP STATISTICS

The Structure of Earnings Survey, which is currently used by Statistics Estonia to calculate the pay gap, the wages survey annex for October, and the combined database created under the pay gap statistics project all have their own positive and negative aspects.

Sample size. The advantage of both the Structure of Earnings Survey and the wages survey annex for October is the extremely large sample size: more than 10,000 companies took part in each, and more than 100,000 individuals participated in the Structure of Earnings Survey and more than 300,000 individuals submitted the annex for October. At the same time, in the combined database of individual surveys, the number of persons used for calculating the pay gap ranges from 12,000 to 15,000, depending on the year.

Economic activities. Considering that the sample of the combined database of individual surveys is rather small compared to that of the Structure of Earnings Survey, the data may be not very reliable for all economic activities. For example, there are not many respondents in the economic activities of mining and quarrying, and electricity, gas, steam and air conditioning supply.

Different background variables. The Structure of Earnings Survey includes a limited number of background variables and the wages survey annex for October allows viewing pay gap only by economic activity. However, the combined database of individual surveys allows viewing pay gap by various background variables (see list in chapter “Pay gap and indicators that affect it”).

Number of hours worked. As in individual surveys, employees themselves are asked how many hours they worked during the survey week, these data are more accurate than in the Structure of Earnings Survey, where the number of hours worked was stated by the company. The hours indicated by the person show how many hours the person actually worked. At the same time, there may be a problem with these data as well if the person worked in several places during the year. Thus, although the total annual earned income is included as earned income, the ‘hours’ variable only takes into account the number of hours worked during the survey week, which may vary in the course of a year.

Resulting from the given positive and negative aspects, it is difficult to determine the right approach for calculating the pay gap. As company-based data collection does not enable retrieving information about employees’ background variables, the only way today is to use individual surveys that contain a variety of background variables that can be used to analyse the pay gap.

Considering that Statistics Estonia continues to move towards using various registers to produce statistics, one potential alternative in the future may be publishing pay gap based on exhaustive register(s) that would give undoubtedly the most accurate results. Unfortunately, registers today lack the sufficient amount of background variables, but there are several registers that are under development and that might in the future allow publishing pay gap along with various background variables.

One possible data source would be the Employment Register (TÖR), which is a register maintained by the Tax and Customs Board and which contains information on employment^a. Considering that the TÖR is currently still under development, it is not possible to publish pay gap based on it right now – this may only be possible in a couple of years from now. If possible, more variables will be added to the TÖR in the future, or it will be linked with other registers, which would make it possible to use TÖR for calculating the gender pay gap by various background variables.

At the moment, Statistics Estonia has ordered the following variables from the TÖR:

- employee’s personal identification code
- employee’s given name
- employee’s surname
- employee’s date of birth
- category of employment, TÖR internal code

^a <http://www.emta.ee/?lang=en&id=35433>

- *occupation, TÖR internal code*
- *code of employing person (commercial registry code or personal identification code)*
- *name of employing company if the code is not a commercial registry code or a personal identification code*
- *commercial registry code of the actual employer*
- *commercial registry code of the person making payments*
- *comment of employing person*
- *code of location of employment*
- *commencement of employment*
- *end of employment*
- *basis for ending employment, TÖR internal code*
- *suspension ID*
- *basis for suspension, TÖR internal code*
- *beginning of suspension period*
- *end of suspension period*
- *initial commencement of employment*
- *time of amendment*
- *time of entry*
- *employment entry status, TÖR internal code*

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
ISIK	Person's unique number in survey				ELFS: Aasta/Kvartal/Leibkond/Kysitletav (Key2) ESS: Leibkond/Haru/Liige (Liikmenr)
STAT_KEY	Unique number of person in Statistics Estonia register				Anonymised personal identification code, stat_unit_id. Persons whose personal identification code was not marked in the survey lack this variable. It is not possible to link ETCB data for those persons.
R_KUU	Survey month		1–12		
UURING	Survey	1 ELFS 2 ESS			
I_SUGU	Sex of person	1 male 2 female			
SYNNIKUU	Birth month of person		1–12		
SYNNIAAST	Birth year of person				
A					
I_VANUS	Age of person as at beginning of survey year				Age of the person as at the beginning of the survey year is calculated as follows: survey year minus birth year minus 1.
I_RAHVUS	Ethnic nationality of person	1 Estonian 2 Russian 3 other 9 unknown			
I_KODAK	Citizenship of person	1 Estonian 2 Russian 3 other 9 undetermined			
I_PEREK_S	Legal marital status of person	1 single, never been married 2 married 3 divorced 4 widowed 9 unknown			
I_PEREK_T	Actual marital status of person	1 single and not cohabiting 2 married and living with husband/wife 3 cohabiting 4 divorced and has not remarried/is not cohabiting 5 widowed and has not remarried/is not cohabiting 6 married and lives apart from legal husband/wife, is not cohabiting 9 unknown			
I_SYNNIRIIK	Birth country of person	1 Estonia 2 other country 9 unknown			
I_ELUK_MK	Code of county of residence	1 Tallinn 37 Harju county 39 Hiiu county 44 Ida-Viru county 49 Jõgeva county 51 Järva county 57 Lääne county 59 Lääne-Viru county 65 Põlva county 67 Pärnu county 70 Rapla county			

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
		74 Saare county 78 Tartu county 82 Valga county 84 Viljandi county 86 Võru county 99 unknown			
I_ELUK_TIH	Place of residence of household by population density	1 high population density 2 medium population density 3 low population density 9 unknown			Level of density is calculated using the code of rural municipality/city of residence. Rules about which rural municipalities and cities belong to which group of population density have been provided. As of 2012, the division is as follows: 1: Tallinn, Tartu, Narva; 2: Elva, Haapsalu, Jõgeva, Jõhvi, Keila, Kiviõli, Kuressaare, Maardu, Paide, Põlva, Pärnu, Rakvere, Saue, Sillamäe, Valga, Viljandi, Võru; 3: other
I_ELUK_LM	Residence of household: city or rural area	1 city 2 rural area 9 unknown			
K_KEEL	Language spoken at home	1 Estonian 2 Russian 3 other 9 unknown			
LAPS_0_3	Number of children up to three years old of the person living in the household of the person				
LAPS_4_6	Number of 4–6-year-old children of the person living in the household of the person				
LAPS_7_17	Number of 7–17-year-old children of the person living in the household of the person				
LAPS_18	Number of children aged 18 years old or older of the person living in the household of the person				
LAPS_ALA	Number of underage (<18) children of the person living in the household of the person				
LAPSI_LK	Total number of children of the person living in the household of the person				
ISCED97M	Highest education level of the person based on ISCED97	1 level 0 2 level 1 3 level 2a 4 level 2c 5 level 3a 6 level 3b 7 level 3c 8 level 4b 9 level 5a 10 level 5b 11 level 6 99 unknown			Questions on education were changed in the surveys in 2008. Therefore, in 2006 and 2007, one set of questions was used, and in 2008–2013 another set was used. This can also be the reason why there are differences in the data.
I_ISCED6	Highest education level of the	0 pre-primary education			

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
	person based on the six levels of ISCED97	1 first stage of basic education 2 second stage of basic education 3 upper secondary education 4 post-secondary non-tertiary education 5 first stage of tertiary education 6 second stage of tertiary education 9 unknown			
I_ISCED3	Highest education level of the person based on the three main levels of ISCED97	1 first level of education: without primary education, primary education, basic education, vocational education without basic education 2 upper secondary education and post-secondary / non-tertiary education: upper secondary education, vocational education/ vocational secondary education/ professional secondary education after basic education, vocational secondary education/vocational education after secondary education 3 Tertiary education: professional secondary education after secondary education, higher education, Master's degree, Doctorate degree 9 unknown			
I_ERIALA_N	Name of field of education attained by person	Classification: Classification of fields of education and training			In the ESS in 2006 and 2007, data on the field of education attained was gathered on a general level only, there is no detailed information.
I_ERIALA_K	Code of field of education attained by person	Classification: Classification of fields of education and training		999	
I_TOOLE	Age of person at the time of commencing continuous employment for the first time	0 has never worked 1–98 age in full years 99 unknown		99	
T_SEIS_2	Socioeconomic status (employed = according to definition, otherwise by personal assessment)	1 employed 2 unemployed 3 studying 4 retired 5 receiving pension for incapacity for work 6 conscript 7 homemaker 8 other inactive			
I_TOOTAV	Person is employed	0 no 1 yes			
I_AMET_N	Description of occupation (Classification ISCO 08)	Classification: ISCO 08			
I_AMET_K	Four-digit code of occupation (Classification ISCO 08)	Classification: ISCO 08		9999	
ISCO1	One-digit code of person's occupation	1 managers 2 professionals 3 technicians and associate professionals 4 clerical support workers 5 service and sales workers 6 skilled agricultural, forestry, hunting and fishery workers 7 craft and related trades workers 8 plant and machine operators, and assemblers			

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
		9 elementary occupations 0 armed forces occupations 99 unknown			
I_TEGEVUS ALA_K	Two-digit economic activity code of the company/body (Classification: EMTAK 2008)	Classification: EMTAK 2008		999	
I_TEGEVUS ALA_N	Economic activity code of the unit of the company/body (Classification: EMTAK 2008)	Classification: EMTAK 2008			
EMTAK1	Alphabetic code of the person’s economic activity	1 A Agriculture, forestry and fishing 2 B Mining and quarrying 3 C Manufacturing 4 D Electricity, gas, steam and air conditioning supply 5 E Water supply; sewerage, waste management and remediation 6 F Construction 7 G Wholesale and retail trade; repair of motor vehicles and motorcycles 8 H Transportation and storage 9 I Accommodation and food service activities 10 J Information and communication 11 K Financial and insurance activities 12 L Real estate activities 13 M Professional, scientific and technical activities 14 N Administrative and support service activities 15 O Public administration and defence; compulsory social security 16 P Education 17 Q Human health and social work activities 18 R Arts, entertainment and recreation 19 S Other service activities 20 T activities of households as employers; undifferentiated goods and services producing activities for households for own use 21 U activities of extraterritorial organisations and bodies 99 unknown			
I_V_TOOLE	Year of commencing employment at last job			9999	
T_AEG	Full- or part-time employment	1 full-time 2 part-time 9 unknown			
T_AEG_imp	Imputed working time format	1 full-time 2 part-time 9 unknown			
t_aeg_i	Working time format has been imputed	0 no 1 yes			
T_TUND	Number of hours usually worked per week			999	EFLS: 1) if person stated their number of hours usually worked per week, this will be the number of hours usually worked per week; 2) if the number of hours

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
					usually worked per week is unknown, the person works full time and their number of hours worked in the previous week falls between 36–60 hours then their number of hours usually worked per week equals the number of hours worked in the previous week; 3) if the number of hours usually worked per week is unknown, the person works part time and the number of hours worked in the previous week falls between 11–35 hours, their number of hours usually worked per week is equal to the number of hours worked in the previous week.
T_TUND_imp	Imputed number of hours usually worked per week				
t_tund_i	Number of hours usually worked per week has been imputed	0 no 1 yes			
T_ASUK_LM	Location of workplace	1 city 2 rural area 3 foreign country 9 unknown			This question was not included in the ESS.
T_ASUK_MK	Code of the county of workplace	1 Tallinn 37 Harju county 39 Hiiu county 44 Ida-Viru county 49 Jõgeva county 51 Järva county 57 Lääne county 59 Lääne-Viru county 65 Põlva county 67 Pärnu county 70 Rapla county 74 Saare county 78 Tartu county 82 Valga county 84 Viljandi county 86 Võru county 97 foreign country 99 unknown			This question was not included in the ESS.
T_ARVGR1	Number of employees in the company/body/unit	1 1 to 10 2 11 to 19 3 20 to 49 4 at least 50 9 unknown			
T_ARV10	Number of employees in a company with up to 10 employees	1 .. 10		99	
T_ARV2GR	Number of employees in company	1 up to 10 2 at least 11 9 unknown			
T_SEISUND	Employment status of person	1 wage employee 2 entrepreneur with wage employee(s), farmer with wage employee(s) 3 self-employed person, farmer without wage employee(s),			

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
		freelancer 4 unpaid worker in a family company, farm 5 other 9 unknown			
T_STAATUS	Status of person (entrepreneur/wage employee)	10 entrepreneur 21 permanent wage employee 22 temporary wage employee 99 unknown			
I_K_TOO	Person has off-hour job	1 yes 2 no 9 unknown			
I_PALGAT	Person is a wage employee	0 no 1 yes			
T_JUHT	Person has managerial responsibilities	1 yes 2 no 9 unknown			
SEKTOR	Sector of the company	1 public sector 2 private sector 9 unknown			Data unavailable in ESS of 2006.
OMAND	Ownership form of company	1 domestically owned 2 foreign-owned 3 other: domestically owned, with foreign investments 9 unknown			Data unavailable in ESS of 2006.
T_LEPING	Type of employment contract	1 written 2 oral 9 unknown			
U_Kpalk_B	Average monthly gross wages in euros stated in the survey (only for wage employees)				Marked only for wage employees. This question wasn't included in the ESS of 2006.
U_Kpalk_i	Monthly gross wages have been imputed in the survey	0 no 1 yes			
Uuringukaal	Expansion factor within the survey				The ELFS and ESS personal weights are used when only one survey is being analysed.
Kaal	Expansion factor of the database				As of the beginning of the survey year, the weight for 14-year-olds is 0.
TT_S1_12_EA	Wages in December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S11_12_EA	Board member remuneration in December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S15_12_EA	Pay received pursuant to a contract in December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
MT_12_EA	Other income in December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing Parental benefit is not

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
					taken into account for other income because in the ETCB database S19 = 0.
TT_S1_01	Wages in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S11_01	Board member remuneration in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S15_01	Pay received pursuant to a contract in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
MT_01	Other income in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_02	Wages in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S11_02	Board member remuneration in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S15_02	Pay received pursuant to a contract in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
MT_02	Other income in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_03	Wages in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S11_03	Board member remuneration in March				0 – has not received this type of income; missing – is

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
					not included in the ETCB database or personal identification code is missing
TT_S15_03	Pay received pursuant to a contract in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
MT_03	Other income in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_04	Wages in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S11_04	Board member remuneration in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
TT_S15_04	Pay received pursuant to a contract in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing
MT_04	Other income in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_05	Wages in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S11_05	Board member remuneration in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_05	Pay received pursuant to a contract in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_05	Other income in May				0 – has not received this

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
					type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_06	Wages in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S11_06	Board member remuneration in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_06	Pay received pursuant to a contract in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_06	Other income in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_07	Wages in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S11_07	Board member remuneration in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_07	Pay received pursuant to a contract in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_07	Other income in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_08	Wages in August				0 – has not received this type of income; missing – is

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
					not included in the ETCB database or personal identification code is missing.
TT_S11_08	Board member remuneration in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_08	Pay received pursuant to a contract in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_08	Other income in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_09	Wages in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S11_09	Board member remuneration in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_09	Pay received pursuant to a contract in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_09	Other income in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_10	Wages in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S11_10	Board member remuneration in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_10	Pay received pursuant to a				0 – has not received this

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
		contract in October			type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_10	Other income in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_11	Wages in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S11_11	Board member remuneration in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_11	Pay received pursuant to a contract in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_11	Other income in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_12	Wages in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S11_12	Board member remuneration in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_12	Pay received pursuant to a contract in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_12	Other income in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
					Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_S1_01_JA	Wages in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S11_01_JA	Board member remuneration in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_S15_01_JA	Pay received pursuant to a contract in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_01_JA	Other income in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
TT_12_EA	Total earned income in December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_01	Total earned income in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_02	Total earned income in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_03	Total earned income in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_04	Total earned income in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_05	Total earned income in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
TT_06	Total earned income in June				<i>missing.</i> 0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_07	Total earned income in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_08	Total earned income in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_09	Total earned income in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_10	Total earned income in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_11	Total earned income in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_12	Total earned income in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_01_JA	Total earned income in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_aasta	Total earned income in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TT_kuude_ar v	The number of months when earned income was received in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TTS1_aasta	Total wages in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
TTS1_kuud	The number of months when wages were received in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TTS11_aasta	Total board member remuneration in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TTS11_kuud	The number of months when board member remuneration was received in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TTS15_aasta	Total pay received pursuant to a contract in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
TTS15_kuud	The number of months when pay was received pursuant to a contract in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
MT_aasta	Total other income in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
MT_kuude_arv	The number of months when other income was received in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_12_EA	Wages paid by the main employer in December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_12_EA	Board member remuneration paid by the main employer for December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_12_EA	Pay received pursuant to a contract from the main employer in December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
PT_MT_12_EA	Other income paid by the main employer in December of the previous year				<i>missing.</i> 0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_01	Wages paid by the main employer in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_01	Board member remuneration paid by the main employer in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_01	Pay received pursuant to a contract from the main employer in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_01	Other income paid by the main employer in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_02	Wages paid by the main employer in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_02	Board member remuneration paid by the main employer in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_02	Pay received pursuant to a contract from the main employer in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_02	Other income paid by the main employer in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
PT_S1_03	Wages paid by the main employer in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_03	Board member remuneration paid by the main employer in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_03	Pay received pursuant to a contract from the main employer in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_03	Other income paid by the main employer in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_04	Wages paid by the main employer in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_04	Board member remuneration paid by the main employer in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_04	Pay received pursuant to a contract from the main employer in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_04	Other income paid by the main employer in April				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_05	Wages paid by the main employer in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_05	Board member remuneration paid by the main employer in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
					<i>missing.</i>
PT_S15_05	Pay received pursuant to a contract from the main employer in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_05	Other income paid by the main employer in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_06	Wages paid by the main employer in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_06	Board member remuneration paid by the main employer in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_06	Pay received pursuant to a contract from the main employer in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_06	Other income paid by the main employer in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_07	Wages paid by the main employer in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_07	Board member remuneration paid by the main employer in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_07	Pay received pursuant to a contract from the main employer in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_07	Other income paid by the main employer in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
					identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_08	Wages paid by the main employer in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_08	Board member remuneration paid by the main employer in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_08	Pay received pursuant to a contract from the main employer in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_08	Other income paid by the main employer in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_09	Wages paid by the main employer in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_09	Board member remuneration paid by the main employer in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_09	Pay received pursuant to a contract from the main employer in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_09	Other income paid by the main employer in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_10	Wages paid by the main employer in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
					<i>missing.</i>
PT_S11_10	Board member remuneration paid by the main employer in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_10	Pay received pursuant to a contract from the main employer in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_10	Other income paid by the main employer in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_11	Wages paid by the main employer in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_11	Board member remuneration paid by the main employer in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_11	Pay received pursuant to a contract from the main employer in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_11	Other income paid by the main employer in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_12	Wages paid by the main employer in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_12	Board member remuneration paid by the main employer in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_12	Pay received pursuant to a contract from the main employer in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
					identification code is missing.
PT_MT_12	Other income paid by the main employer in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_S1_01_J A	Wages paid by the main employer in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S11_01_ JA	Board member remuneration paid by the main employer in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_S15_01_ JA	Pay received pursuant to a contract from the main employer in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_01_J A	Other income paid by the main employer in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_12_EA	Total earned income paid by the main employer in December of the previous year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_01	Total earned income paid by the main employer in January				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_02	Total earned income paid by the main employer in February				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_03	Total earned income paid by the main employer in March				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_04	Total earned income paid by the main employer in April				0 – has not received this type of income; missing – is not included in the ETCB

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
					database or personal identification code is missing.
PT_05	Total earned income paid by the main employer in May				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_06	Total earned income paid by the main employer in June				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_07	Total earned income paid by the main employer in July				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_08	Total earned income paid by the main employer in August				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_09	Total earned income paid by the main employer in September				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_10	Total earned income paid by the main employer in October				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_11	Total earned income paid by the main employer in November				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_12	Total earned income paid by the main employer in December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_01_JA	Total earned income paid by the main employer in January of the following year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_aasta	Total earned income paid by the main employer in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_kuude_ar v	Number of months when earned income was received from the main employer in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
					identification code is missing.
PTS1_aasta	Total wages paid by the main employer in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PTS1_kuud	Number of months when wages were received from the main employer in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PTS11_aasta	Total board member remuneration from the main employer in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PTS11_kuud	Number of months when board member remuneration was received from the main employer in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PTS15_aasta	Total pay received pursuant to a contract from the main employer in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PTS15_kuud	Number of months when pay was received pursuant to a contract from the main employer in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
PT_MT_aasta	Total other income from the main employer in a year: January–December				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing. Parental benefit is not taken into account for other income because in the ETCB database S19 = 0.
PT_MT_kuud	Number of months when other income was received from the main employer in a year				0 – has not received this type of income; missing – is not included in the ETCB database or personal identification code is missing.
ADDITIONAL VARIABLES					
T_ja_TT	According to a survey, the person is employed and has, according to the ETCB, received earned income	0 no 1 yes			
PT_ja_TT	According to a survey, the person is a wage employee and has, according to the ETCB, received earned income	0 no 1 yes			
VAN_GR1	Age groups	1 24 or younger 2 25–34 3 35–44			

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
		4 45–54 5 55–64 6 65–74 7 75+			
Eriala_gr	Field of highest level of education attained	0 no professional training 1 teacher training and education science 2 the humanities and arts 3 social sciences, business and law 4 natural and exact sciences 5 technology, manufacture and construction 6 agriculture, forestry and fisheries, veterinary science 7 human health and well-being 8 service 9 unknown			0 – employed but does not have professional training
P_suhe	Person is in a relationship	0 no 1 yes			Not in relationship = single or has never been married; widowed or divorced, has not remarried or is not cohabiting; in a relationship = is married/cohabiting. Persons whose actual marital status is unknown have been classified as “not in relationship”.
TT_ak	Earned income: average per month in the calendar year				Total earned income in a calendar year is divided by the number of months when earned income was received.
TT_uk	Earned income: in survey month				Earned income in survey month.
TT_ujk	Earned income: in the month following survey month				Earned income in the month following the survey month.
TT_3k	Earned income: the average of the month preceding survey month, survey month and the month following survey month				Earned income in the month preceding survey month, survey month and the month following survey month, divided by the number of months when earned income was received in that period.
PT_ak	Earned income paid by the main employer: average per month in a calendar year				Total earned income of a calendar year divided by the number of months when earned income was received.
PT_uk	Earned income paid by the main employer: in survey month				Earned income in survey month.
PT_ujk	Earned income paid by the main employer: in month following survey month				Earned income in month following the survey month.
PT_3k	Earned income paid by the main employer: average of month preceding survey month, survey month and month following survey month				Earned income in the month preceding the survey month, survey month and the month following the survey month divided by the number of months when earned income was received.
TT_tund	Total earned income, average per hour				Formula for calculating: (monthly average earned

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values “Does not know” and “Refuses to answer”	Comment
					income per year * 12) / ((hours worked per week / 5) * number of working days in the year). The number of working days in a year has been determined as 253 which is the average of the years 2006–2014. The average has been used so that wages would not change when there are more or fewer working days in a year.
EMTAK10	10 main groups of EMTAK	1 A Agriculture, forestry and fishing 2 B–E Industry 3 F Construction 4 G–I Trade, transportation, accommodation and food service activities 5 J Information and communication 6 K Financial and insurance activities 7 L Real estate activities 8 M–N Professional, scientific and technical activities, administrative and support service activities 9 O–Q Public administration, education, human health and social work activities 10 R–U Arts, entertainment and recreation, other service activities, other 99 unknown			Division is based on the NACE 10 division published in Eurostat where: A – Agriculture, forestry and fishing B–E – Industry (except construction) F – Construction G–I – Wholesale and retail trade, transport, accommodation and food service activities J – Information and communication K – Financial and insurance activities L – Real estate activities M–N – Professional, scientific and technical activities; administrative and support service activities O–Q – Public administration, defence, education, human health and social work activities R–U – Arts, entertainment and recreation; other service activities; activities of household and extra-territorial organizations and bodies
laps_EK	Number of preschool age children (0–6 years old)	0 no preschool age children present 1 1 preschool age child 2 at least 2 preschool age children			
Juht	Managers by ISCO and managerial responsibilities	0 no 1 yes			If ISCO1 code = 1 and person has responded that he/she has managerial responsibilities.
Mees_LT	Person is male and job is classified as ‘elementary occupation’	0 no 1 yes			Following the main conclusion of the analysis of the University of Tartu: male wage employees whose job, by the Classification of Occupations, is classified as ‘elementary occupation’ and whose company operates in the economic activity of manufacturing based on EMTAK.
T_tund_s	Prescribed working hours			999	Calculated as follows: part

Annex 1. Description of the combined database

Name of variable	Explanation of variable	Explanation of values	Range of values	Values "Does not know" and "Refuses to answer"	Comment
					time and usually works less than 35 hours: the number of hours usually worked according to the estimate of the respondent; part time and usually works at least 35 hours: 20 hours (the average number of hours of part-time employees); full time and works in the field of education (two-digit code of occupation = 23): 35 hours; full time and works elsewhere: 40 hours.
<i>TT_tund_s</i>	Total average hourly earned income according to the number of prescribed working hours				Formula for calculating: (monthly average earned income in a year * 12) / ((hours worked per week / 5) * number of working days in the year). The number of working days in a year has been determined as 253 which is the average of the years 2006–2014. The average has been used so that wages would not change when there are more or fewer working days in a year.
<i>Ettev</i>	Person is entrepreneur	0 no 1 yes			Persons whose status is entrepreneur (<i>T_STAATUS</i> =10).
<i>UP_tund</i>	Hourly gross wages based on the wages stated in the survey				
<i>UP_tund_s</i>	Hourly wages based on the gross wages stated in the survey (by prescribed working hours)				
<i>PT_tund</i>	Hourly wages paid by main employer by the number of working hours stated in the survey				
<i>PT_tund_s</i>	Hourly wages paid by main employer by prescribed working hours				

Annex 2 (I). Ranges of part-time and full-time employees' average hourly gross wages by economic activity, 2006–2009
(euros)

	2006		2007		2008		2009									
	Part-time	Full-time	Part-time	Full-time	Part-time	Full-time	Part-time	Full-time								
	-30%	+30%	-30%	+30%	-30%	+30%	-30%	+30%								
Average of economic activities	1.95	3.62	2.52	4.68	2.35	4.36	3.05	5.67	2.78	5.16	3.51	6.53	2.88	5.35	3.40	6.32
Agriculture, forestry and fishing	1.69	3.13	2.04	3.78	2.03	3.77	2.47	4.58	2.41	4.47	2.84	5.28	2.15	3.98	2.65	4.92
Mining and quarrying	2.02	3.76	2.89	5.37	2.43	4.52	3.50	6.51	2.89	5.36	4.03	7.49	3.01	5.60	3.98	7.40
Manufacturing	1.71	3.17	2.35	4.37	2.05	3.81	2.85	5.30	2.43	4.51	3.29	6.10	2.92	5.43	3.22	5.97
Electricity, gas, steam and air conditioning supply	1.81	3.37	2.93	5.44	2.18	4.05	3.55	6.59	2.58	4.80	4.09	7.59	2.94	5.47	4.40	8.18
Water supply; sewerage, waste management and remediation	1.70	3.15	2.56	4.76	2.04	3.79	3.11	5.77	2.42	4.49	3.58	6.64	2.36	4.38	3.45	6.41
Construction	1.98	3.67	2.77	5.15	2.38	4.41	3.36	6.24	2.82	5.23	3.87	7.18	3.12	5.79	3.47	6.44
Wholesale and retail trade; repair of motor vehicles and motorcycles	1.59	2.96	2.44	4.53	1.92	3.56	2.96	5.49	2.27	4.22	3.40	6.32	2.39	4.43	3.19	5.92
Transportation and storage	1.75	3.25	2.47	4.58	2.11	3.91	2.99	5.55	2.50	4.64	3.44	6.39	2.65	4.92	3.28	6.09
Accommodation and food service activities	1.48	2.74	1.70	3.15	1.77	3.29	2.06	3.82	2.10	3.91	2.37	4.40	2.15	3.99	2.32	4.31
Information and communication	2.50	4.64	3.95	7.34	3.01	5.58	4.79	8.90	3.56	6.62	5.52	10.25	3.63	6.74	5.49	10.20
Financial and insurance activities	2.65	4.92	4.03	7.48	3.19	5.92	4.88	9.06	3.78	7.02	5.62	10.43	4.27	7.94	5.65	10.50
Real estate activities	1.67	3.10	1.97	3.66	2.01	3.73	2.39	4.43	2.38	4.42	2.75	5.10	2.59	4.82	2.77	5.14
Professional, scientific and technical activities	2.18	4.05	2.93	5.44	2.62	4.87	3.55	6.60	3.11	5.78	4.09	7.60	3.39	6.30	3.93	7.29
Administrative and support service activities	1.47	2.73	2.26	4.20	1.77	3.29	2.74	5.10	2.10	3.90	3.16	5.87	2.17	4.03	3.13	5.81
Public administration and defence; compulsory social security	2.08	3.86	3.11	5.78	2.50	4.64	3.77	7.00	2.96	5.50	4.34	8.07	3.11	5.77	4.13	7.68
Education	2.22	4.13	2.23	4.15	2.67	4.96	2.71	5.02	3.17	5.88	3.12	5.79	3.13	5.81	3.10	5.77
Human health and social work activities	2.78	5.16	2.48	4.60	3.34	6.20	3.00	5.57	3.96	7.36	3.45	6.42	3.98	7.40	3.43	6.38
Arts, entertainment and recreation	1.56	2.90	1.96	3.64	1.88	3.48	2.37	4.41	2.22	4.13	2.73	5.08	2.18	4.05	2.62	4.86
Other service activities	1.41	2.63	1.87	3.48	1.70	3.16	2.27	4.22	2.02	3.74	2.62	4.86	2.03	3.77	2.51	4.65

Annex 2 (II). Ranges of part-time and full-time employees' average hourly gross wages by economic activity, 2010–2013 (euros)

	2010		2011		2012		2013									
	Part-time	Full-time	Part-time	Full-time	Part-time	Full-time	Part-time	Full-time								
	-30%	+30%	-30%	+30%	-30%	+30%	-30%	+30%								
Average of economic activities	2.82	5.25	3.39	6.30	2.91	5.40	3.58	6.64	2.98	5.54	3.79	7.05	3.20	5.94	4.07	7.57
Agriculture, forestry and fishing	2.15	3.99	2.79	5.18	2.32	4.30	2.80	5.20	2.42	4.50	3.07	5.69	2.58	4.78	3.57	6.63
Mining and quarrying	2.90	5.38	4.20	7.80	3.47	6.44	4.54	8.44	3.56	6.62	4.70	8.72	3.84	7.14	5.19	9.65
Manufacturing	2.66	4.95	3.23	6.01	2.81	5.23	3.39	6.29	2.84	5.28	3.69	6.85	3.07	5.69	3.97	7.37
Electricity, gas, steam and air conditioning supply	2.89	5.37	4.51	8.38	3.33	6.18	4.84	9.00	3.49	6.47	5.40	10.04	3.69	6.85	5.84	10.84
Water supply; sewerage, waste management and remediation	2.18	4.05	3.41	6.32	1.93	3.58	3.49	6.49	2.07	3.85	3.84	7.14	2.25	4.19	4.11	7.63
Construction	2.74	5.09	3.49	6.49	3.22	5.98	3.62	6.72	3.19	5.93	4.00	7.42	3.18	5.90	4.28	7.94
Wholesale and retail trade; repair of motor vehicles and motorcycles	2.37	4.41	3.10	5.75	2.41	4.47	3.42	6.34	2.52	4.68	3.60	6.68	2.66	4.94	3.77	7.01
Transportation and storage	2.72	5.05	3.33	6.19	2.32	4.30	3.40	6.31	2.56	4.75	3.45	6.41	2.79	5.17	3.86	7.16
Accommodation and food service activities	2.01	3.74	2.19	4.07	2.16	4.00	2.32	4.32	2.23	4.15	2.47	4.59	2.21	4.11	2.63	4.88
Information and communication	3.71	6.90	5.49	10.19	3.82	7.10	5.97	11.09	3.84	7.14	6.19	11.49	4.38	8.14	6.67	12.39
Financial and insurance activities	4.77	8.87	5.50	10.21	5.42	10.06	5.71	10.61	5.82	10.80	5.90	10.96	6.18	11.48	6.44	11.96
Real estate activities	2.09	3.87	2.35	4.36	2.33	4.33	2.67	4.95	2.39	4.43	2.84	5.27	2.54	4.72	2.98	5.54
Professional, scientific and technical activities	3.80	7.05	4.16	7.73	3.54	6.57	4.38	8.13	3.93	7.29	4.87	9.04	3.86	7.18	4.85	9.01
Administrative and support service activities	2.19	4.07	3.16	5.87	2.16	4.00	3.30	6.12	2.11	3.91	3.44	6.38	2.37	4.39	3.74	6.94
Public administration and defence; compulsory social security	3.14	5.83	4.05	7.52	3.19	5.92	4.15	7.71	3.49	6.49	4.34	8.06	3.76	6.98	4.75	8.81
Education	3.06	5.68	3.09	5.74	3.10	5.76	3.16	5.86	3.19	5.93	3.26	6.06	3.51	6.51	3.53	6.55
Human health and social work activities	3.82	7.10	3.28	6.09	4.07	7.55	3.44	6.38	4.07	7.55	3.63	6.73	4.44	8.24	3.90	7.24
Arts, entertainment and recreation	2.57	4.78	2.68	4.98	2.26	4.20	2.91	5.40	2.28	4.23	3.21	5.95	2.44	4.52	3.07	5.71
Other service activities	1.84	3.42	2.28	4.23	1.86	3.45	2.06	3.82	1.79	3.33	2.21	4.10	1.82	3.38	2.37	4.39

Annex 3. Gender pay gap by Eurostat, Statistics Estonia and combined database, 2013
(per cent)

Economic activity	Eurostat	Statistics Estonia	Combined database
Economic activities total: A–S	Does not publish	24.8	22.6
Economic activities B–S, except O	29.9 ^a	Does not publish	25.7
Economic activities B–S, incl. O	28.2	Does not publish	23.5
A. Agriculture, forestry and fishing	Does not publish	12.7	20.7
B. Mining and quarrying	34.3	33.7	37.0
C. Manufacturing	32.8	31.5	32.8
D. Electricity, gas, steam and air conditioning supply	19.6	18.1	19.4
E. Water supply; sewerage, waste management and remediation	9.2	13.6	35.5
F. Construction	19.4	21.3	17.1
G. Wholesale and retail trade; repair of motor vehicles and motorcycles	39.4	33.0	33.9
H. Transportation and storage	6.7	–0.2	4.5
I. Accommodation and food service activities	24.7	19.3	21.8
J. Information and communication	31.6	30.1	28.5
K. Financial and insurance activities	44.9	41.8	45.7
L. Real estate activities	22.9	10.5	32.6
M. Professional, scientific and technical activities	18.5	21.4	16.6
N. Administrative and support service activities	17.4	18.4	7.0
O. Public administration and defence; compulsory social security	9.8	12.1	7.2
P. Education	25.5	26.4	29.6
Q. Human health and social work activities	28.5	22.0	29.8
R. Arts, entertainment and recreation	28.4	20.3	14.9
S. Other service activities	16.5	27.4	29.6

^a Only the economic activities B–S have been taken into account to measure and monitor the pay gap indicator used in the European Union, the economic activity O has not been included. Companies with fewer than 10 employees have not been included either.

Sources: Statistics Estonia, Eurostat^b

^b Indicators of Statistics Estonia: Table WS5335: Gender Pay Gap by Economic Activity (EMTAK 2008) October of the database of Statistics Estonia; Indicators of Eurostat: Table Gender pay gap in unadjusted form in % – NACE Rev. 2 (structure of earnings survey methodology) of the database of Eurostat (earn_gr_gpgr2).