Overview of the project

Country: Estonia

Type of research: Randomized Controlled Trial

Title: "Motivating Improved Healthcare Using Holistic Patient Contracts"

Study period: June 2021 - March 2023

Authors:

- Kevin Croke (Harvard T.H. Chan School of Public Health)
- Benjamin Daniels (Georgetown University)
- Daniel Rogger (World Bank)
- Robert Lipinski (World Bank; underline corresponding author for the reproducibility package)

The code in this replication package constructs analysis file for the "Motivating Improved Healthcare Using Holistic Patient Contracts" paper manuscript (Croke, Daniels, Lipinski and Rogger 2024).

All scripts are written in R programming language. The code should run in approx. 12 hours. Twenty scripts are submitted.

The scripts produce[1]:

- 4 main tables;

- 8 appendix tables;

- 2 main body figures (each composed of two sub-figures);

- 2 appendix figures (each composed of two sub-figures).

---

[1] Tables and Figures not created by the code were prepared manually (Figure 1 and Figure A1;) or extracted from pre-existing documentation (ECM care plans in Section A1; Table A1). More details on those below.

Project background

As **Estonia** was undergoing a democratic transition from the Soviet era, it also began to transform its health model away from the prevailing "Semashko" model of centralized, hospital-focused care to a more primary care-centered approach. The current model is funded through a social health insurance system, and primary care is provided by independent family physicians who contract directly with the **Estonia Health Insurance Fund** (EHIF). The shift in model allocates greater responsibility for the quality of healthcare services to independent physicians.

All Estonians covered by EHIF are assigned to a private family physician. The approximately 800 family physicians who provide primary care in Estonia are paid through a mix of capitation fees (66%), allowances (16%), and fee-for-service (20%). To motivate providers to provide quality services as determined by the Estonian Health Insurance Fund, a small performance-based element is included in doctor payments called the **Quality Bonus System** (QBS). Though it accounts for a relatively small amount (2-4%) of total provider compensation, it is now a well-established element of the primary care system, and participation has been compulsory since 2015. The initial goal of the QBS system was to signal to family physicians that, in a new family medicine system of primary care, it was their responsibility to focus on improving preventive care and management of chronic disease.

The Estonian Health Insurance Fund (EHIF) maintains electronic health records (EHR) describing every "episode of care" for all of its covered individuals since 2009. Since EHIF is a payer and not a care provider, these records are organized as **billing claims** records and do not have qualitatively detailed case histories. Each service provided for a given patient at a given provider for a given episode is recorded as a line entry in a "**procedures**" dataset and is together identifiable as an episode of care by "bill numbers." Bill numbers uniquely identify an episode of care between a single provider and patient, and they close when the provider requests reimbursement for the episode. The procedures provided during an episode of care are linked by their bill number to one or more entries in a corresponding "**diagnosis**" dataset, which provides the ICD-10 code(s) for the condition(s) which the services in the episode were intended to diagnose and/or treat.

The claims and diagnosis datasets of EHR are each divided into **eight types of care**: day care services, inpatient services, inpatient nursing services, inpatient rehabilitation services, outpatient services, outpatient rehabilitation services, outpatient nursing services, and primary healthcare services. In addition to the billed services, the claims records include metadata for each episode of care. This includes (as relevant) the type of arrival[2], duration of treatment, type of specialty doctor involved, and type of hospital. Each of the eight corresponding claims datasets across contain both a procedures dataset and a diagnosis dataset. They can therefore be used to identify whether a given condition is new or ongoing.

---

[2] For instance, walk-in, doctor referral, ambulance transport

The Enhanced Care Management (ECM) RCT consists of training and coaching family physicians and their teams to develop holistic care and proactive outreach plans for chronically ill patients or those vulnerable to developing chronic illnesses. The core goal of ECM is to improve the quality of care provided to complex patients, including by increasing the use of preventive care, improving coordination of care across health system levels, and increasing patient involvement in care. These elements can improve patient health and quality of life and may reduce the need for curative medical services—for example, by supporting patients with type 2 diabetes to improve their diet and increase physical activity to limit further deterioration in their health.

ECM practices include improved tracking of tests and referrals, follow-up by PHC providers after hospital discharges, tracking of medication adherence, monitoring of patients between clinic visits, and greater focus on clinical quality. It includes four elements: identifying high-risk patients through risk stratification, developing care management plans by the primary care physician in consultation with the patient, proactively linking care providers together, and developing a team approach between patients and their caregivers. ECM reflects global primary care reforms that aim to focus the health system's attention on high-risk groups and improve the continuity of care for these patients (Peikes, 2018).

A pilot of the program was first conducted in 2017 with 10 providers[3], focused on patients with multiple chronic conditions, including cardiovascular disease (CVD), hypertension, diabetes, and elevated blood lipids, and other conditions. Evaluation of the pilot showed that providers made 40% more calls to patients; were 11% more likely to have patients on appropriate statin prescriptions; had patients 25% less likely to be hospitalized for CVD-related conditions; and were 11% more likely to follow up within 30 days in the event of an acute CVD incident (Kurowski et al., 2015). However, this pilot was conducted with a purposively selected group of 10 doctors who were expected to be highly motivated early adopters, limiting the possibility of inference about the causal impact of the program or its likely effectiveness at scale. A larger and more rigorous study was therefore warranted to test whether the impressive results achieved in the pilot can be sustained at scale.

The following project implements a scaled-up version of ECM programme, targeting 93 providers, of whom 72 ended up participating in the programme. The study randomized patients at those providers into treatment (N=1,781) and control (N=3,275) conditions. The health outcomes of those patients are measured after 22 months after the intervention started.

The manuscript accompanying this README file contains further information on the methodology, variables, and models used.

---

[3] 'Providers' is a term used interchangeably with 'GP' to refer to family physician.

Data Availability and Provenance Statements

All the healthcare billing micro-data files used to measure to health-related outcomes of patients in the program were shared with the researchers by the Estonian Health Insurance Fund (hereafter EHIF) – the public national insurance organization in Estonia. The files were shared under a Data License Agreement, signed on 12th Nov 2019 between World Bank DIME and EHIF. The datasets were shared for the purposes of **internal WB replication. <u>No part of patient-level data can be made public without prior agreement from EHIF</u>**, which remains the owner of the data. Hence, no data files are included in the public reproducibility package.

Data availability statement

All the results in the paper use confidential microdata from the billing records of the EHIF. To gain access to the microdata, EHIF personnel needs to be contacted under info@tervisekassa.ee (also see: https://www.tervisekassa.ee/andmeparingud). Data License Agreement needs to be set and signed before any microdata can be shared outside the organization.

Statement about Rights

- ☑ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

- ☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Summary of Availability

- ☐ All data **are** publicly available.

- ☐ Some data **cannot be made** publicly available.

- ☑ **No data can be made** publicly available.

Details on each Data Source

Patient-level healthcare billing micro-data are not provided in a single file, but rather are divided up along three key dimensions. The files, as provided originally by EHIF, are all stored in relevant sub-folders in 'Data/Raw' folder, as explained below.

First of all, files contain different parts of the **healthcare bill information,** i.e. there are files that only contain the overall **bill information** (e.g. start and end date, type of care) and the files that specify the 'contents' of each bill, i.e. either the **diagnoses** or the medical **procedures** claimed under each bill. The specific variables to be found in those three different types of files, referred to as **billing**, **diagnoses** and **procedure** files are specified in the table below. This 3-way distinction is also used to **order the sub-folders** in both 'Data/Raw' and 'Data/Clean' sub-folders**.** Those different types of data are also described in section A3 of the Appendix in the paper.

Second, all three types of files are also divided, as per original versions sent by EHIF, according to the **type of healthcare** claimed under a bill into 8 distinctive types:

1. Primary healthcare
2. Outpatient care
3. Outpatient rehabilitation
4. Outpatient nursing
5. Inpatient care
6. Inpatient rehabilitation
7. Inpatient nursing
8. Day care

Third, the files come from different **periods**, as specified in the file names, corresponding to the different batches of data shared with us by EHIF. Those two latter dimensions are incorporated into file names, but not into the folder structure.

A separate category of healthcare data are **prescriptions**. Those are not directly related and cannot be matched to other billing data. They also have their own structure, as specified in the table below. The only dividing dimension for prescription files is the time range they capture.

Table 1: Raw Variables Description

Data type	Variable name	Description	Used in code?
	billnr	Unique ID number of a healthcare bill	Yes
	typeoftreatment	Type of treatment, almost perfectly corresponding to the type of the file as either primary healthcare, outpatient care, day care etc.	Yes
Billing	patientidencrypted	Unique ID number of a patient	Yes
	residencecodesettlementlevel	Place of residence	No
	residencecodelocalgovlevel	Local government (LG) of residence, as code	No
	residencenamelocalgovlevel	Local government (LG) of residence, as name	No

	gender	Gender of a patient	Yes
	dateofbirth	Date of birth of a patient[4]	Yes
	dateofdeath	Date of death of a patient	Yes
	age	Age of a patient at the time of the start of a treatment	No
	codeofdischargetype	The reasons for patient discharge [EHIF][5]	Yes
	codeofadmissiontypeoradmitt	The method of a patient's admission into treatment, as code [EHIF]	Yes
	admissiontypeoradmittingdoct	The method of a patient's admission into treatment, as name [EHIF]	No
	startoftreatment	Start of the billing/episode of care	Yes
	endoftreatment	End of the billing/episode of care	Yes
	codeofhealthcareprovider	Clinic where billing/episode of care occurred, as code	No
	nameofhealthcareprovider	Clinic where billing/episode of care occurred, as name	No
	codeofhospitaltype	Type of hospital, as code	No
	nameofhospitaltype	Type of hospital, as name	No
	doctorid	Unique ID number of a doctor	No
	codeofdoctorsspeciality	Speciality of the doctor issuing the billing, as code	No
	nameofdoctorsspeaciality	Speciality of the doctor issuing the billing, as name	No
	specialtyascontractedbythee	Type of service, as code	No
	nameofspecialtyascontractedbythee	Type of service, as name	No
	drgcode	Drug code	No
	mdccode	Drug category	No
Diagnoses	billnr	Unique ID number of a healthcare bill	Yes
	codeofdiagnos	Diagnosis, as ICD-10 code (see Data/Clean/Other/Diagnosis_code_desc.csv)	Yes
	typeofdiagnos	Whether diagnosis primary (Põhidiagnoos) or secondary (Kaasuv) with + (primary ilness); - (recurrent disease); 0 (diagnosis unconfirmed)	No (for now)
	numberofdiagnosis	Severity of the diagnosis	No (for now)
Procedures	billnr	Unique ID number of a healthcare bill	Yes
	procedure	Procedure performed, as code (see Data/Raw/Other/costs_procedures.csv for exact names)	Yes
	dateofprocedure	Date procedure performed	Yes
	Nrooftimes	Number of times procedure performed, as specified in EHIF billing guidelines	Yes
Prescriptions	patientidencrypted	Unique ID number of a patient	Yes
	healthcarefacilityid	Prescribing healthcare facility, as code	No
	healthcarefacilityname	Prescribing healthcare facility, as name	Ni
	doctorid	Unique ID number of a doctor	No
	specialtyofdoctor	Speciality of the doctor issuing the billing, as name	No
	prescription_status	What happened with a prescription[6]	Yes
	dateofprescription	Date prescription issues	Yes
	dateofpurchase	Date prescription purchased	Yes
	prescription_dgn	Diagnosis associated with a prescription	No
	atccode	ATC code of the drug prescribed[7]	Yes

---

[4] All raw dates across all the files are specified in yyyymmdd format, e.g. 16th November 2023 becomes 20231116.

[5] Explanation of the codes for the variables labelled with [EHIF] can be found under: https://www.riigiteataja.ee/akt/125112011004, which is also hyperlinked in the table

[6] See Data/Clean/Other/Codebook_prescriptions.pdf

activesubstance	Active substance in the drug prescribed	No
codeofpackaging	Driug prescribed, as code	No
nameofpackaging	Driug prescribed, as name	No
pricepaidbyehif	Total cost of a prescription that was paid by EHIF	Yes
totalpriceofprescription	Total cost of a prescription	Yes
numberofpackaging	Number of packages/units prescribed	No

Finally, the Data/Raw folder contains '**ECM Inclusion'** and '**Other'** sub-folders. The former contains all the files necessary to identify patient groups, principally clinics, providers, and patients assigned to ECM treatment and control conditions. The latter sub-folder contains any other files used, that didn't meet the criteria of other sub-folders and that are not patient-level micro-data. Details are provided below.

The 'Data/Clean' folder parallels 'Data/Raw' in its structure. The main difference are two additional folders stored there:

- 'Aggregated' folder stores combined and grouped datasets created by scripts 3 through 5.
- 'Re-randomizations' folder stores re-randomizations of treatment assignment used to generate Table A8 and Figure A3 and prepared by script 'A08_tableA8_robustness.R'

'Data/Clean' folder is mostly populated only by running the code. The code also adds .parquet copies of the raw .csv files in Billing, Diagnoses, Prescriptions, and Procedures sub-folders of 'Data/Raw'. This is done to minimize re-run times as .parquet files are read considerably quicker. Again, those are not provided as a part of the replication package (but can be added if requested), but rather are created by running the code ('02_bills_clean.R'). So, after running the code 'Data/Raw' should contain .parquet version of each billing, diagnosis, procedures, and prescriptions file, while 'Data/Clean' should contained i) clean ii) combined iii) aggregated version of each of those.

The exact cleaning operations and aggregation procedures are specified in the notes withing the individual R scripts.

---

[7] ATC stands for Anatomical Therapeutic Chemical code and is the international drug classification scheme maintained by the WHO. See the drug with their associated ATC codes at:
https://www.ravimiregister.ee/Default.aspx?pv=HumRavimid.ATCPuu&ot=C&l=en#C

Table 2: List of datasets

Data name	Data files	Location	Provider
Billing	Data/Raw/Billing/day_care_2009_2019_10.csv Data/Raw/Billing/day_care_2022_23.csv Data/Raw/Billing/daycare_2019_2022_append.csv Data/Raw/Billing/inpatient_care_2009_2019_10.csv Data/Raw/Billing/inpatient_care_2019_2022_append.csv Data/Raw/Billing/inpatient_care_2021_2022.csv Data/Raw/Billing/inpatient_care_2022.csv Data/Raw/Billing/inpatient_care_2022_23.csv Data/Raw/Billing/inpatient_nursing_care_2009_2019_10.csv Data/Raw/Billing/inpatient_nursing_care_2019_2022_append.csv Data/Raw/Billing/inpatient_nursing_care_2022_23.csv Data/Raw/Billing/inpatient_rehabilitation_2009_2019_10.csv Data/Raw/Billing/inpatient_rehabilitation_2019_2022_append.csv Data/Raw/Billing/inpatient_rehabilitation_2022_23.csv Data/Raw/Billing/outpatient_2009_2019_10.csv Data/Raw/Billing/outpatient_2022_23.csv Data/Raw/Billing/outpatient_care_2019_2022_append.csv Data/Raw/Billing/outpatient_nursing_2019_2022_append.csv Data/Raw/Billing/outpatient_nursing_care_2009_2019_10.csv Data/Raw/Billing/outpatient_nursing_care_2022_23.csv Data/Raw/Billing/outpatient_rehabilitation_2009_2019_10.csv Data/Raw/Billing/outpatient_rehabilitation_2019_2022_append.csv Data/Raw/Billing/outpatient_rehabilitation_2022_23.csv Data/Raw/Billing/primaryhealthcare_2009_2019_10.csv Data/Raw/Billing/primaryhealthcare_2019_2022_append.csv Data/Raw/Billing/primaryhealthcare_2022_23.csv	Data/Raw/Billing	EHIF
Diagnoses	Data/Raw/Diagnoses/day_care_diag_2009_2019_10_old.csv Data/Raw/Diagnoses/day_care_diag_2019_10_new.csv Data/Raw/Diagnoses/day_care_diag_2021_2022.csv Data/Raw/Diagnoses/day_care_diag_2022.csv Data/Raw/Diagnoses/day_care_diag_2022_23.csv Data/Raw/Diagnoses/daycare_diag_2019_2020_2021.csv Data/Raw/Diagnoses/daycare_diag_2019_2022_append.csv Data/Raw/Diagnoses/daycare_diag_2022.csv Data/Raw/Diagnoses/inpatient_care_diag_2009_2019_10_old.csv Data/Raw/Diagnoses/inpatient_care_diag_2019_10_new.csv Data/Raw/Diagnoses/inpatient_care_diag_2019_2020_2021.csv Data/Raw/Diagnoses/inpatient_care_diag_2019_2022_append.csv Data/Raw/Diagnoses/inpatient_care_diag_2021_2022.csv Data/Raw/Diagnoses/inpatient_care_diag_2022.csv Data/Raw/Diagnoses/inpatient_care_diag_2022_23.csv Data/Raw/Diagnoses/inpatient_diag_2019_2022_append.csv Data/Raw/Diagnoses/inpatient_nursing_care_diag_2009_2019_10_old.csv Data/Raw/Diagnoses/inpatient_nursing_care_diag_2019_10_new.csv Data/Raw/Diagnoses/inpatient_nursing_care_diag_2019_2020_2021.csv Data/Raw/Diagnoses/inpatient_nursing_care_diag_2019_2022_append.csv	Data/Raw/Diagnoses	EHIF

Data/Raw/Diagnoses/inpatient_nursing_care_diag_2021_2022.csv		
Data/Raw/Diagnoses/inpatient_nursing_care_diag_2022.csv		
Data/Raw/Diagnoses/inpatient_nursing_care_diag_2022_23.csv		
Data/Raw/Diagnoses/inpatient_rehabilitation_diag_2009_2019_10_old.csv		
Data/Raw/Diagnoses/inpatient_rehabilitation_diag_2019_10_new.csv		
Data/Raw/Diagnoses/inpatient_rehabilitation_diag_2019_2020_2021.csv		
Data/Raw/Diagnoses/inpatient_rehabilitation_diag_2019_2022_append.csv		
Data/Raw/Diagnoses/inpatient_rehabilitation_diag_2021_2022.csv		
Data/Raw/Diagnoses/inpatient_rehabilitation_diag_2022.csv		
Data/Raw/Diagnoses/inpatient_rehabilitation_diag_2022_23.csv		
Data/Raw/Diagnoses/outpatient_diag_2009_2019_10_old.csv		
Data/Raw/Diagnoses/outpatient_diag_2019_10_new.csv		
Data/Raw/Diagnoses/outpatient_diag_2019_2020_2021.csv		
Data/Raw/Diagnoses/outpatient_diag_2019_2022_append.csv		
Data/Raw/Diagnoses/outpatient_diag_2021_2022.csv		
Data/Raw/Diagnoses/outpatient_diag_2022 (1).csv		
Data/Raw/Diagnoses/outpatient_diag_2022 (2).csv		
Data/Raw/Diagnoses/outpatient_diag_2022 (3).csv		
Data/Raw/Diagnoses/outpatient_diag_2022 (4).csv		
Data/Raw/Diagnoses/outpatient_diag_2022 (5).csv		
Data/Raw/Diagnoses/outpatient_diag_2022 (6).csv		
Data/Raw/Diagnoses/outpatient_diag_2022 (7).csv		
Data/Raw/Diagnoses/outpatient_diag_2022.csv		
Data/Raw/Diagnoses/outpatient_diag_2022_23.csv		
Data/Raw/Diagnoses/outpatient_nursing_care_diag_2009_2019_10_old.csv		
Data/Raw/Diagnoses/outpatient_nursing_care_diag_2019_10_new.csv		
Data/Raw/Diagnoses/outpatient_nursing_care_diag_2019_2020_2021.csv		
Data/Raw/Diagnoses/outpatient_nursing_care_diag_2019_2022_append.csv		
Data/Raw/Diagnoses/outpatient_nursing_care_diag_2021_2022.csv		
Data/Raw/Diagnoses/outpatient_nursing_care_diag_2022.csv		
Data/Raw/Diagnoses/outpatient_nursing_care_diag_2022_23.csv		
Data/Raw/Diagnoses/outpatient_rehabilitation_diag_2009_2019_10_old.csv		
Data/Raw/Diagnoses/outpatient_rehabilitation_diag_2019_10_new.csv		
Data/Raw/Diagnoses/outpatient_rehabilitation_diag_2019_2020_2021.csv		
Data/Raw/Diagnoses/outpatient_rehabilitation_diag_2019_2022_append.csv		
Data/Raw/Diagnoses/outpatient_rehabilitation_diag_2021_2022.csv		
Data/Raw/Diagnoses/outpatient_rehabilitation_diag_2022.csv		
Data/Raw/Diagnoses/outpatient_rehabilitation_diag_2022_23.csv		
Data/Raw/Diagnoses/primaryhealthcare_diag_2009_2019_10_old.csv		
Data/Raw/Diagnoses/primaryhealthcare_diag_2019_2020_2021.csv		
Data/Raw/Diagnoses/primaryhealthcare_diag_2019_2022_append.csv		
Data/Raw/Diagnoses/primaryhealthcare_diag_2021_2022.csv		
Data/Raw/Diagnoses/primaryhealthcare_diag_2022.csv		
Data/Raw/Diagnoses/primaryhealthcare_diag_2022_23.csv		

Procedures	Data/Raw/Procedures/day_care_procedures_2009_2019_10.csv Data/Raw/Procedures/day_care_procedures_2022_23.csv Data/Raw/Procedures/daycare_procedures_2019_2022_append.csv Data/Raw/Procedures/inpatient_care_procedures_2009_2019_10.csv Data/Raw/Procedures/inpatient_care_procedures_2019_2022_append.csv Data/Raw/Procedures/inpatient_care_procedures_2021_2022.csv Data/Raw/Procedures/inpatient_care_procedures_2022 (1).csv Data/Raw/Procedures/inpatient_care_procedures_2022.csv Data/Raw/Procedures/inpatient_care_procedures_2022_23.csv Data/Raw/Procedures/inpatient_nursing_care_procedures_2009_2019_10.csv Data/Raw/Procedures/inpatient_nursing_care_procedures_2019_2022_append.csv Data/Raw/Procedures/inpatient_nursing_care_procedures_2022_23.csv Data/Raw/Procedures/inpatient_rehabilitation_procedures_2009_2019_10.csv Data/Raw/Procedures/inpatient_rehabilitation_procedures_2019_2022_append.csv Data/Raw/Procedures/inpatient_rehabilitation_procedures_2022_23.csv Data/Raw/Procedures/outpatient_nursing_care_procedures_2009_2019_10.csv Data/Raw/Procedures/outpatient_nursing_care_procedures_2022_23.csv Data/Raw/Procedures/outpatient_nursing_procedures_2019_2022_append.csv Data/Raw/Procedures/outpatient_procedures_2009_2019_10.csv Data/Raw/Procedures/outpatient_procedures_2019_2022_append.csv Data/Raw/Procedures/outpatient_procedures_2022_23.csv Data/Raw/Procedures/outpatient_rehabilitation_procedures_2009_2019_10.csv Data/Raw/Procedures/outpatient_rehabilitation_procedures_2019_2022_append.csv Data/Raw/Procedures/outpatient_rehabilitation_procedures_2022_23.csv Data/Raw/Procedures/primaryhealthcare_procedures_2009_2019_10.csv Data/Raw/Procedures/primaryhealthcare_procedures_2019_2022_append.csv Data/Raw/Procedures/primaryhealthcare_procedures_2022_23.csv	Data/Raw/Procedures	EHIF
Prescriptions	C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2009_2013.csv C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2014_2017.csv C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2018_2019_10.csv C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2019_2020.csv C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2020.csv C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2021.csv C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2022_1.csv C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2022_2.csv	Data/Raw/Prescriptions	EHIF

	C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2022_23.csv C:/Users/ASUS/Documents/World_Bank/Locker/Estonia/Health/ECM (replication)/Data/Raw/Prescriptions/prescriptions_2022_3.csv		
ECM Inclusion	clinic_list_112020.csv Clinics.csv ECM_randomization_results_1_bothID.csv ECM_randomization_results_2_bothID.csv Patient_ID_1.csv Patient_ID_2.csv	Data/Raw/ECM Inclusion	
ECM Inclusion	ECM_eligible_may21.csv ECM_eligible_nov22.csv GP_final_participation_list.csv QBS Need Adjustment Calculations.csv stratified blocks.csv treat_control_clinic_gp.csv	Data/Raw/ECM Inclusion	All files were constructed and/or shared by EHIF specifically for the purposes of this RCT.
Other	costs_procedures.csv costs_procedures (translated).csv	Data/Raw/Other	Dataset recording prices for all medical procedures, with a temporal breakdown, are obtained from EHIF website. Translated version was obtained using Google Translator
Other	ECM Evaluation of the care plans ECM Evaluation of the care plans-b	Data/Raw/Other	Data gathered as part of the external evaluators surveying of randomly selected healthcare plans created as part of ECM. See details in section A3.4 of the paper for details

Reproduction instructions

1. Open the RStudio project file "RR_EST_2024_101.Rproj" in RStudio.

2. Run the command "renv::restore()" to reproduce the project environment and install all required dependencies. This step is required only the first time the reproducibility package runs.

3. Open the script "000_MASTER.R" and run it.

Computational requirements

Software Requirements

Analyses were generated using R version 4.2.2. File "`00_global.R`" sets global parameters, defines custom functions, and loads dependencies required. It needs be run prior as the first file from "`000_MASTER.R`". The following packages are used in the analysis.

- NCmisc 1.2.0
- survminer 0.4.9
- ggpubr 0.6.0
- broom 1.0.5
- collapse 2.0.7
- DescTools 0.99.49
- wildrwolf 0.7.0
- pdftools 3.3.3
- survival 3.4-0
- patchwork 1.1.2
- vip 0.3.2
- ivtools 2.3.0
- psych 2.3.3
- ggtext 0.1.2
- estimatr 1.0.0
- knitr 1.43
- multiwayvcov 1.2.3
- lmtest 0.9-40
- zoo 1.8-12
- fixest 0.11.2
- tidylog 1.0.2
- plm 2.6-3
- Rmpfr 0.9-4
- gmp 0.7-3
- scales 1.3.0
- arrow 12.0.0
- car 3.1-2
- carData 3.0-5
- foreign 0.8-84
- stringi 1.7.8
- data.table 1.14.6
- janitor 2.2.0
- haven 2.5.1
- lubridate 1.9.2
- forcats 1.0.0
- stringr 1.5.1
- dplyr 1.1.2
- purrr 1.0.1
- readr 2.1.4

- `tidyr 1.3.0`
- `tibble 3.2.1`
- `ggplot2 3.4.4`
- `tidyverse 2.0.0`

Controlled Randomness

Random seed generated using random.org, on 29/11/2023 11:02:26 UTC, from numbers between 1 and 10^8

- ☑ Random seed is set at line **56** of program '`00_global.R`'

Memory and Runtime Requirements

The code was last run on a 8-core AMD Ryzen-based laptop with Windows 11 Home 64-bit operating system (version 23H2)

Approximate time needed to reproduce the analyses on a standard 2024 desktop machine:

- ☐ <10 minutes

- ☐ 10-60 minutes

- ☐ 1-2 hours

- ☐ 2-8 hours

- ☐ 8-24 hours

- ☑ 1-3 days

- ☐ 3-14 days

- ☐ > 14 days

- ☐ Not feasible to run on a desktop machine, as described below.

Most scripts should run in <5 minutes. The bulk of the total time is spent on 6 scripts:

- `02_bills_clean.R` – takes around 2-3 hours to run.; combines all micro data-files of different type and date (see explanation below)
- `03_outcomes_diagnosis.R` – takes approx. 30 minutes to run; prepares all diagnosis-level micro-data
- `04_outcomes_prescriptions.R` – takes approx.. 15 minutes to run; prepares all prescription-level micro-data
- `05_outcomes_procedures.R` – takes approx.. 60 minutes to run prepares all procedure-level micro-data

- **A08_tableA8_robustness.R** – takes approx. 6-8 hours to run; prepares randomized inference and other robustness checks for Table A9
- **A09_figureA3_robustness.R** – takes approx. 12-24 hours to run; creates randomization inference survival curves and p-values shown in Figure A3.

Description of programs/code

000_MASTER.R

Purpose: *Sources (run) all other scripts from one file*

00_global.R

Estimated run time: *<1 minute*

Input files:

- Data/Clean/Other/Outcome_dict.xlsx

Output files:

- *None*

Purpose:

- Load all packages
- Set global parameters (XXX)
- Create custom functions
- Read dictionary file

Notes:

- Please note that you might need to adjust the main project path to run the code

01_clean_id.R

Estimated run time: *1-2 minutes*

Input files:

- Data/Raw/ECM Inclusion/Patient_ID_1.csv
- Data/Raw/ECM Inclusion/Patient_ID_2.csv

- Data/Raw/ECM Inclusion/Clinics.csv
- Data/Raw/ECM Inclusion/clinic_list_112020.csv

- Data/Raw/ECM Inclusion/treat_control_clinic_gp.csv
- Data/Raw/ECM Inclusion/GP_final_participation_list.csv

- Data/Raw/ECM Inclusion/QBS Need Adjustment Calculations.csv

- Data/Raw/ECM Inclusion/ECM_randomization_results_1_bothID.csv
- Data/Raw/ECM Inclusion/ECM_randomization_results_2_bothID.csv

- Data/Raw/ECM Inclusion/ECM_eligible_may21.csv

- Data/Raw/ECM Inclusion/ECM_eligible_nov22.csv

Output files:

- Data/Clean/ECM Inclusion/id_codes.csv
- Data/Clean/ECM Inclusion/clinic_all.csv
- Data/Clean/ECM Inclusion/list_all.csv
- Data/Clean/ECM Inclusion/patient_ecm_randomized.csv
- Data/Clean/ECM Inclusion/patient_eligible1_all.csv
- Data/Clean/ECM Inclusion/patient_eligible2_no_excluded.csv
- Data/Clean/ECM Inclusion/patient_eligible3_lists.csv
- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv
- Data/Clean/ECM Inclusion/patient_eligible4_pure_control.csv
- Data/Clean/ECM Inclusion/patient_eligible5_ecm_clinics.csv
- Data/Clean/ECM Inclusion/patient_eligible6_control_treat_lists.csv
- Data/Clean/ECM Inclusion/patient_eligible7_control_treat_patients.csv
- Data/Clean/ECM Inclusion/patient_eligible8_participating.csv

Purpose:

- Prepare full patient ID lists
- Create treatment group assignment at the patient-level: (i) ECM treatment group patients; (ii) ECM control group patients; (iii) all patients meeting ECM inclusion criteria (the 'pure control' group).
- Add the required clinic- and provider-level variables for each patient
- Calculate numbers used in randomization plots shown in Figure 1 and A1

Notes:

- Figure A1 provides a graphic overview of the cleaning and sub-setting operations performed in this script

02_bills_clean.R

Estimated run time: ~60 minutes

Input files:

- *All .csv OR .parquet files[8] in Data/Raw/*
  - *Billing*
  - *Diagnoses*
  - *Raw*

Output files:

- *All of the above saved as .parquet rather than .csv in the original folders*
- *All of the above in Data/Clean rather than Data/Raw*

- Data/Clean/Billing/billing_all.parquet
- Data/Clean/Diagnoses/diagnoses_all.parquet
- Data/Clean/Procedures/procedures_all.parquet

Purpose:

- Row-bind datasets from different types of healthcare and periods (see 'Data/Raw' in 'Folder architecture' section)
- Remove bills with dates falling outside of the observation period
- Remove duplicates
- Remove patients not in ECM or pure control group
- Remove unnecessary variables
- Extract patient-level demographic data (gender and date of birth)

Notes:

- The structure of the files differs between files coding different parts of the healthcare bills, i.e. billing, diagnoses, procedures, and prescriptions (as per **Table 1**)**.** However, this structure is constant for each of those data categories for all time periods and all 8 healthcare types. This is why, the main purpose of this script is to row-bind all those datasets into one aggregate dataset per billing, diagnoses and procedures files, after doing some basic cleaning, the most important of which involves sub-setting data which is originally capturing whole of Estonian population, to just the relevant groups of patients from the selected clinics (see 01_clean_id.R and Figures 1 and A1)

03_outcomes_diagnosis.R

Estimated run time: 10-15 minutes

Input files:

- Data/Clean/Diagnoses/diagnoses_all.parquet
- Data/Clean/Other/ ICD codes of acute conditions.xlsx
- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv

Output files:

---

[8] Once .parquet files are created, they are equivalent (but faster to work with) to .csv files.

- Data/Clean/ECM Inclusion/deaths_all.csv
- Data/Clean/Diagnoses/dta_diagnosis.parquet
- Data/Clean/Aggregated/Diagnoses_outcomes_month_18_23.parquet
- Data/Clean/Aggregated/Diagnoses_outcomes_year_18_23.parquet
- Data/Clean/Aggregated/Diagnoses_outcomes_year_rel_18_23.parquet
- Data/Clean/Aggregated/Diagnoses_outcomes_period_18_23.parquet

Purpose:

- Extract dates of death per patient
- Create all outcome variables related to diagnosis
- Group and summarize diagnosis-related outcomes at patient-level by:
  - Month
  - Year (calendar)
  - Year (relative to ECM onset in June 2021)
  - Pre-/post-treatment period

04_outcomes_prescriptions.R

Estimated run time: 20-30 minutes

Input files:

- *All of the above saved as .parquet[9] rather than .csv in the original folders*
- *All of the above in Data/Clean rather than Data/Raw*

Output files:

- *All of the above in Data/Clean rather than Data/Raw and with saved as .parquet rather than .csv*

- Data/Clean/Prescriptions/prescriptions_all.parquet
- Data/Clean/Prescriptions/dta_prescriptions.parquet

- Data/Clean/Aggregated/Prescriptions_outcomes_month_18_23.parquet
- Data/Clean/Aggregated/Prescriptions_outcomes_year_18_23.parquet
- Data/Clean/Aggregated/Prescriptions_outcomes_year_rel_08_23.parquet
- Data/Clean/Aggregated/Prescriptions_outcomes_period_18_23.parquet

Purpose:

- Clean and merge prescription files
- Create all outcome variables related to prescriptions
- Group and summarize prescription-related outcomes at patient-level by:
  - Month
  - Year (calendar)
  - Year (relative to ECM onset in June 2021)

---

[9] Once .parquet files are created, they are equivalent (but faster to work with) to .csv files.

    o Pre-/post-treatment period

05_outcomes_procedures.R

Estimated run time: 60 minutes

Input files:

- Data/Clean/Procedures/procedures_all.parquet
- Data/Clean/Other/ICD codes of acute conditions.xlsx
- Data/Clean/ECM Inclusion/patient_ecm_eligible_demo.csv

Output files:

- Data/Clean/ECM Inclusion/patient_ecm_accept.csv
- Data/Clean/Procedures/dta_procedures.parquet
- Data/Clean/Aggregated/Procedures_outcomes_month_18_23.parquet
- Data/Clean/Aggregated/Procedures_outcomes_year_18_23.parquet
- Data/Clean/Aggregated/Procedures_outcomes_year_rel_18_23.parquet
- Data/Clean/Aggregated/Procedures_outcomes_period_18_23.parquet

Purpose:

- Extract dates of enrolment/inclusion into ECM (based on the special code for ECM enrolment visit)
- Create outcome variables related to procedures
- Group and summarize procedure-related outcomes at patient-level by:
  - Month
  - Year (calendar)
  - Year (relative to ECM onset in June 2021)
  - Pre-/post-treatment period

06_table1_balance_patient.R

Estimated run time: *2-3 minutes*

Input files:

- Data/Clean/Aggregated/All_outcomes_period_18_23.parquet
- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv

Output files:

- Tables/CSV/all/table1_balance_patient_pre.csv

- *LaTeX code for the balance table to copy to Overleaf*

Purpose:

- Create Table1 - pre-treatment balance table comparing pure control, ECM control, and ECM treatment groups across all key outcomes.

Notes:

- Code can be adjusted to create separate tables for all patients; those classified as mild risk; and those classified as high risk patients

07_table2_and_A56_cross.R

Estimated run time: *2-3 minutes*

Input files:

- Data/Clean/Aggregated/All_outcomes_period_18_23.parquet
- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv

Output files:

- Tables/CSV/all/table2_cross.csv
- Tables/CSV/mild/table2_cross.csv
- Tables/CSV/severe/table2_cross.csv

- *LaTeX code for the balance tables to copy to Overleaf*

Purpose:

- Create cross-sectional and ANCOVA results table comparing pure control, ECM control, and ECM treatment groups across all outcomes.

Notes:

- Code creates three equivalent versions of the table for i) all ii) mild-risk iii) severe-risk patients

08_table3_hosp_deaths.R

Estimated run time: *2-3 minutes*

Input files:

- Data/Clean/Diagnoses/dta_diagnosis.parquet
- Data/Clean/ECM Inclusion/patient_eligible7_control_treat_patients.csv
- Data/Clean/ECM Inclusion/patient_ecm_accept.csv

Output files:

- Data/Aggregated/dta_deaths_temp.RData
- Tables/CSV/table3_hosp_deathsa.csv
- Tables/CSV/table3_hosp_deathsb.csv
- Tables/CSV/table3_hosp_deathsc.csv
- Tables/CSV/table3_hosp_deathsd.csv

- *LaTeX code for the balance table to copy to Overleaf*

Purpose:

- Create mortality regression results for ECM treatment and control groups – both OLS and Cox Proportional-Hazards Models with and without 'treatment x risk group' interaction term
- Calculate life expectancy across patient groupings
- Prepare data frame ready for survival analysis and survival curve plotting in the next two scripts

09_table4_hosp_deaths_sub.R

Estimated run time: *2-3 minutes*

Input files:

- Data/Clean/Aggregated/dta_deaths_temp.RData

Output files:

- Tables/CSV/table3_hosp_deaths_suba.csv
- Tables/CSV/table3_hosp_deaths_subb.csv
- Tables/CSV/table3_hosp_deaths_subc.csv
- Tables/CSV/table3_hosp_deaths_subd.csv

- Data/Clean/dta_deaths_temp.rds

- *LaTeX code for the balance table to copy to Overleaf*

Purpose:

- Create mortality regression results for ECM treatment and control groups – both OLS and Cox Proportional-Hazards Models – separately for mild-risk and severe-risk patients

10_figures23_surv_curves.R

Estimated run time: *1-2 minutes*

Input files:

- Data/Clean/Aggregated/dta_deaths_temp.RData

Output files:

- Figures/Figure 2 Survival Hospitalization (mild).png
- Figures/Figure 2 Survival Hospitalization (severe).png
- Figures/Figure 3 Survival Mortality (mild).png
- Figures/Figure 3 Survival Mortality (severe).png

Purpose:

- Create survival curves for time until first hospitalization and time until death – separately for mild-risk and severe-risk patients

A01_figureA2_rand_blocs.R

Estimated run time: *<1 minute*

Input files:

- Data/Clean/ECM Inclusion/clinic_all.csv
- Data/Raw/ECM Inclusion/stratified blocks.csv
- Data/Clean/ECM Inclusion/patient_eligible7_control_treat_patients.csv

Output files:

- Figures/Figure A2a. Randomization (clinic).png
- Figures/Figure A2b. Randomization (patient).png

Purpose:

- Show randomization and stratification outcomes visually at both the clinic- and GP-level

A02_tableA2_codebook.R

Estimated run time: *<1 minute*

Input files:

- Data/Clean/Other/Outcome_dict.csv

Output files:

- *LaTeX code for the balance table to copy to Overleaf*

Purpose:

- Re-code variable dictionary stored as .csv into LaTeX table code to include as a codebook in Table A2

A03_tableA3_clinic.R

Estimated run time: *<1 minute*

Input files:

- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv
- Data/Clean/Raw/ECM Evaluation of the care plans.csv
- Data/Clean/Raw/ECM Evaluation of the care plans-b.csv

Output files:

- Tables/CSV/tableA3_clinic.csv
- Data/Clean/Other/Care plan evaluations.csv

- *LaTeX code for the balance table to copy to Overleaf*

Purpose:

- Create merged file with survey-based evaluation of individual ECM care plans
- Create clinic- and GP-level balance table

A06_tableA6_interact.R

Estimated run time: *3-4 minutes*

Input files:

- Data/Clean/Aggregated/All_outcomes_period_18_23.parquet
- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv
- Data/Clean/Other/Care plan evaluations.csv

Output files:

- Tables/as_csv/all/tableA6_interact.csv

- *LaTeX code for the balance table to copy to Overleaf*

Purpose:

- Create table of results from interaction models
- Interactions terms included:
    - ECM treatment X QBS (provider)
    - ECM treatment X Management quality (provider)
    - ECM treatment X Care plan evaluation (provider)
    - ECM treatment X Past outcomes (patient)

Notes:

- Code can be adjusted to create separate tables for all patients; those classified as mild risk; and those classified as high risk patients

A07_tableA7_IV.R

Estimated run time: *<1 minute*

Input files:

- Data/Clean/Aggregated/All_outcomes_period_18_23.parquet
- Data/Clean/ECM Inclusion/patient_ecm_accept.csv
- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv
- Data/Clean/ECM Inclusion/patient_eligible7_control_treat_patients.csv

Output files:

- Tables/CSV /all/tableA7_IV.csv

- *LaTeX code for the balance table to copy to Overleaf*

Purpose:

- Create intent-to-treat (instrumental variable) regression results comparing patients assigned AND enrolling into ECM with ECM control group.

Notes:

- Code can be adjusted to create separate tables for all patients; those classified as mild risk; and those classified as high risk patients

A08_tableA8_robustness.R

Estimated run time: *6-8 hours*

Input files:

- Data/Clean/Aggregated/All_outcomes_period_18_23.parquet
- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv

Output files:

- Data/Clean/Re-randomizations/randp10000 (seed 9559388).Rdata
- Tables/CSV/all/tableA8_robustness_10000.csv
- Data/Clean/randp_reg_10000_count.csv
- Data/Clean/randp_reg_10000_dummy.csv

- *LaTeX code for the balance table to copy to Overleaf*

Purpose:

- Re-randomize treatment assignment in 10,000 ways, saving the output to i) avoid re-estimation at every run ii) use in 'A09_figureA3_robustness.R'
- Create table comparing p-values for models presented in Table 2, as re-estimated using:
  - Benjamini-Hochberg procedure
  - Romano-Wolf multiple hypothesis correction
  - Randomized inference (10,000 iterations)

A09_figureA3_robustness.R

Estimated run time: *12-24 hours*

Input files:

- Data/Clean/Aggregated/All_outcomes_period_18_23.parquet
- Data/Clean/Diagnoses/dta_diagnosis.parquet
- Data/Clean/ECM Inclusion/patient_eligible4_eligible_clinics.csv
- Data/Re-randomizations/randp10000 (seed 9559388).Rdata

Output files:

- Figures/Figure A3 Survival Hospitalization_10000_9559388.png
- Figures/Figure A3 Survival Mortality_10000_9559388.png
- Tables/figureA3_randp_death_all.csv
- Tables/figureA3_randp_death_mild.csv
- Tables/figureA3_randp_death_severe.csv
- Tables/figureA3_randp_hosp_all.csv
- Tables/figureA3_randp_ hosp_mild.csv
- Tables/figureA3_randp_ hosp_severe.csv

Purpose:

- Visually compare the survival rates (overall and until the first hospitalization) for all, mild-risk, and severe-risk patients as actually recorded in the data against 10,000 possible re-randomizations
- Calculate randomization inference p-values shown in the notes below the figures

List of tables and programs

The provided code reproduces:

- ☐ All numbers provided in text in the paper

- ☑ All tables and figures in the paper (*apart from Table A1; see below*)

- ☐ Selected tables and figures in the paper, as explained and justified below.

Figure/Table #	Program	Output file	Notes
Table 1	06_table1_balance_patient.R	Tables/CSV/all/table1_balance_patient_pre.csv	
Table 2	07_table2_and_A56_cross.R	Tables/CSV/all/table2_cross.csv	
Table 3	08_table3_hosp_deaths.R	Tables/CSV/all/table3_hosp_deathsa.csv Tables/CSV/all/table3_hosp_deathsb.csv Tables/CSV/all/table3_hosp_deathsc.csv Tables/CSV/all/table3_hosp_deathsd.csv	Panels of the table are saved separately and then combined in the code to produce LaTeX-ready otuput
Table 4	09_table4_hosp_deaths_sub.R	Tables/CSV/all/table3_hosp_deaths_suba.csv Tables/CSV/all/table3_hosp_deaths_subb.csv Tables/CSV/all/table3_hosp_deaths_subc.csv Tables/CSV/all/table3_hosp_deaths_subd.csv	Panels of the table are saved separately and then combined in the code to produce LaTeX-ready otuput
Table A1	-	-	Manually re-created based on: WB (2018) Revising Estonia's Quality Bonus Scheme in Primary Care" (see 'Documentation' folder)
Table A2	A03_tableA3_clinic.R	Tables/CSV/tableA3_clinic.csv	
Table A3	A02_tableA2_codebook.R	Data/Clean/Outcome_dict.csv	The table is almost equivalent to the Outcome_dict.csv – file that serves as variable dictionary throughout the analyses; the code only re-orders it and cleans it up for

			immediate copy-pasting into LaTeX.
Table A4	07_table2_and_A56_cross.R	Tables/CSV/mild/table2_cross.csv	
Table A5	07_table2_and_A56_cross.R	Tables/CSV/severe/table2_cross.csv	
Table A6	A06_tableA6_interact.R	Tables/CSV/all/tableA6_interact.csv	
Table A7	A07_tableA7_IV.R	Tables/CSV/all/tableA8_IV.csv	
Table A8	A08_tableA8_robustness.R	Tables/CSV/all/tableA8_robustness10000.csv	
Figure 1	01_id_clean.R	No output file;	Numbers copied manually from R Studio to Word version of the Figure
Figure 2	10_figures23_surv_curves.R	Figure 2 Survival Hospitalization (mild).png Figure 2 Survival Hospitalization (severe).png	

Figure 3	10_figures23_surv_curves.R	Figure 3 Survival Mortality (mild).png Figure 3 Survival Mortality (severe).png	
Figure A1	01_id_clean.R	No output file;	Numbers copied manually from R Studio to Word version of the Figure
Figure A2	A01_figureA2_rand_blocs.R	Figure A2a. Randomization (clinic).png Figure A2b. Randomization (patient)	
Figure A3	A09_figureA3_robustness.R	Figure A3 Survival Hospitalization_10000_9559388 Figure A3 Survival Mortality_10000_9559388.png Tables/figureA3_randp_death_all.csv Tables/figureA3_randp_death_mild.csv Tables/figureA3_randp_death_severe.csv Tables/figureA3_randp_hosp_all.csv Tables/figureA3_randp_ hosp_mild.csv Tables/figureA3_randp_ hosp_severe.csv	