

Supplemental Appendix

Defying Distance?

The Provision of Medical Services in the Digital Age

Amanda Dahlstrand

February 12, 2026

Appendix Table of Contents

A1	Appendix tables	2
A2	Appendix figures	24
B	Appendix: Variables	31
C	Representativeness	34
1	Representativeness of the patient sample	34
2	Comparison of sample doctors to nationwide in-person primary care	35
D	Additional results and background information	36
1	In-person primary care: Socioeconomic inequality	36
1.1	How patients choose in-person primary care clinics	37
2	Doctors' time use and patients' satisfaction with doctors	37
3	Informational requirements for reallocations	38
E	Data appendix	38
1	Definition of the analysis sample	38
2	Important variables	40
3	Data sources	43
F	Robustness	45
1	Robustness of risk variables and complementarity functional form	45
1.1	Nonparametric investigation of the functional form of the skill-risk interaction	45
1.2	Robustness of risk variable choice	46
1.3	Semiparametric complementarity using LASSO-predicted risk	47
2	Robustness exercise for both value added and complementarity	48
3	Robustness on doctor volume	49
G	Estimation of the distribution of doctor skills	49
H	Tests of the value-added method	51
1	Forecast bias	51
2	Autocorrelations	52

A1 Appendix tables

Table A1: Diagnosis code frequencies in online versus in-person primary care consultations

	Online (%)	In-person (%)
Respiratory system	18.0	15.5
Symptoms	14.7	16.9
Health status factors	13.8	8.5
Skin & subcutaneous tissue	13.5	7.0
Infectious	12.0	5.2
Genitourinary system	7.0	6.3
Injury, poisoning	5.3	7.1
Eye and adnexa	4.6	1.6
Musculoskelet., connective	2.7	7.0
Mental and behavioral	2.6	8.5
Digestive system	1.8	3.1
Nervous system	1.4	1.3
Ear and mastoid process	1.4	5.5
Neoplasms	0.4	1.2
Circulatory system	0.2	1.5
Pregnancy, childbirth	0.2	1.3
Diseases of the blood	0.2	0.4
Endocrine, nutrit., metabolic	0.2	2.0
Observations	213,866	44,992
Unique patients	213,866	18,661

Note: This table shows the frequency of diagnosis (ICD code) categories in the analysis sample and for a comparable sample in in-person primary care. The in-person sample consists of patients observed in the analysis sample *and* in in-person primary care in Region Scania during 2015. This year is chosen to avoid overlap with the analysis sample period of 2016–2018, when patients may also have had on-line care. The in-person column shows ICD code category frequencies among these patients’ in-person primary care visits in Region Scania during 2015. Symptoms refers to ”Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99),” and Health status factors refers to ”Factors influencing health status and contact with health services (Z00-Z99).” The latter is further explained as circumstances that cannot be classifiable to categories A00-Y89 in the ICD 10. In the interest of space, I have omitted four ICD code categories with frequencies lower than 0.2% in both online and in-person consultations.

Table A2: Sample restrictions

	Patients	Doctors
All visits	631,664	769
+ Include only patient's first visit with the service	377,769	716
+ Visit took place between 2016 and 2018	376,270	701
+ Patient found in Statistics Sweden's registry	376,270	701
+ "Drop-in" consultations	309,777	515
+ No pediatricians	302,639	500
+ No small children	233,470	499
+ Jointly defined controls	233,296	499
+ 1 month end of sample truncation	219,454	487
+ Doctor observed over at least two two-week periods	218,977	435
+ Remove each doctor's first two-week period	211,883	435

Note: This table shows how adding sample restrictions affects the number of observed patients and doctors. The Statistics Sweden registry which one row refers to contains patient demographic and socioeconomic data, and the row shows that all patients are matched to this registry, even though there can be missing values for individual variables. "Drop-in consultations" are those for which there is a first-come-first-served allocation of patients to doctors, with the following exceptions: I exclude pediatricians and small children under the age of five because they are not always assigned in a first-come-first-served manner. The control characteristics I require to be defined for all consultations are the basic ones: the count of the patient's earlier diagnosed Elixhauser-related comorbidities, age, and gender. I exclude the final month of the data to ensure that all patient outcomes can be accurately measured over the 30 day outcome period. The final two rows reflect the restriction to only using doctors' past data when estimating their skill for future (two-week) time periods.

Table A3: Interaction effects between doctor skill and patient risk for high-frequency time periods

	Dependent variable ($\times 100$)		
	Counter-Guideline	Primary Care	Emergency Room
Doctor skill (std) ($\hat{\beta}_1^k$)	-0.92 (0.21)	-1.03 (0.23)	-0.19 (0.08)
Patient risk (std) ($\hat{\beta}_2^k$)	0.15 (0.03)	11.76 (0.43)	3.22 (0.15)
Doctor skill $\times$ Patient risk ($\hat{\beta}_3^k$)	-0.12 (0.06)	-0.45 (0.45)	-0.56 (0.16)
Constant	1.72 (0.03)	25.11 (0.11)	5.55 (0.06)
Date-time shift FE	✓	✓	✓
Observations	161,135	161,135	161,135
Outcome mean	1.69	25.1	5.53
Mean for low risk	1.64	22.3	4.82
Mean for high risk	2.07	66.8	13.6

Note: This table shows estimates of complementarities between doctor skill and patient risk following Equation 12 but with the added sample restriction that each observed doctor $\times$ two-week time-period cell must contain at least 30 observations. This table is to be compared to the main Table 6. For a description of the outcomes and patient risk measures, see Sections II.C and II.D, respectively. Patient risks and doctor skills are standardized at the doctor $\times$ two-week time-period level. High-risk patients, shown in the bottom rows reporting sample means, are defined as those with risks at least 1 s.d. above the mean within each two-week period; the rest are considered low risk. Standard errors are bootstrapped with 1000 iterations.

Table A4: Interaction effects between skill and risk for high-volume doctors

	Dependent variable ($\times 100$)		
	Counter-Guideline	Primary Care	Emergency Room
Doctor skill (std) ($\hat{\beta}_1^k$)	-0.88 (0.05)	-0.93 (0.20)	-0.16 (0.07)
Patient risk (std) ($\hat{\beta}_2^k$)	0.18 (0.03)	11.59 (0.40)	3.29 (0.14)
Doctor skill $\times$ Patient risk ($\hat{\beta}_3^k$)	-0.094 (0.047)	-0.15 (0.41)	-0.52 (0.16)
Constant	1.77 (0.03)	25.12 (0.09)	5.60 (0.05)
Date-time shift FE	✓	✓	✓
Observations	198,060	198,060	198,060
Outcome mean	1.73	25.1	5.56
Mean for low risk	1.67	22.3	4.85
Mean for high risk	2.17	67.2	13.7

Note: This table shows estimates of complementarities between doctor skill and patient risk following Equation 12 but with the added sample restriction that each observed doctor must have at least the median number of observations. This table is to be compared to the main Table 6. The analysis sample of 435 doctors is restricted here to the 218 doctors who have weakly above the median number of within-sample consultations; the median is 206 consultations. For a description of the outcomes and patient risk measures, see Sections II.C and II.D, respectively. Patient risks and doctor skills are standardized at the doctor $\times$ two-week time-period level. High-risk patients, shown in the bottom rows reporting sample means, are defined as those with risks at least 1 s.d. above the mean within each two-week period; the rest are considered low risk. Standard errors are bootstrapped with 1000 iterations.

Table A5: Doctor-patient balance tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	General Practitioner	Other specialty	Medical Doctor only	In specialist training	Completed specialist training	Lives in Stockholm Region	Speaks non-EU language	Top medical school graduation	Years of experience	Total consultations in all data (1000s)	Female doctor
Number of comorbidities	-0.0017 (0.0011)	0.0019 (0.0011)	-0.00012 (0.0011)	0.00046 (0.0011)	-0.00035 (0.0011)	-0.00093 (0.0012)	-0.00099 (0.0012)	-0.00052 (0.00085)	-0.0059 (0.0099)	-0.0045 (0.0082)	-0.0017 (0.0013)
Female patient	-0.0013 (0.0021)	-0.00026 (0.0020)	0.0016 (0.0021)	0.0031 (0.0021)	-0.0047 (0.0020)	0.0022 (0.0022)	0.0013 (0.0022)	-0.00085 (0.0016)	-0.00033 (0.018)	0.020 (0.015)	0.0043 (0.0025)
Patient age	0.00019 (0.000091)	0.0000092 (0.000088)	-0.00020 (0.000091)	0.000044 (0.000092)	0.00016 (0.000087)	0.000010 (0.000096)	-0.000061 (0.000094)	0.000077 (0.000069)	-0.00079 (0.00080)	-0.00090 (0.00066)	-0.000072 (0.00011)
1 st generation immigrant	-0.00027 (0.0032)	-0.0010 (0.0031)	0.0013 (0.0033)	-0.0021 (0.0033)	0.00079 (0.0031)	-0.00088 (0.0034)	0.0035 (0.0034)	-0.0019 (0.0024)	-0.011 (0.028)	-0.019 (0.024)	0.0039 (0.0039)
2 nd generation immigrant	-0.0021 (0.0041)	-0.00028 (0.0040)	0.0023 (0.0041)	-0.0070 (0.0041)	0.0046 (0.0039)	-0.0019 (0.0043)	0.0011 (0.0042)	0.0036 (0.0031)	0.022 (0.036)	0.0050 (0.030)	0.0058 (0.0048)
Lives in Stockholm Region	0.0029 (0.0020)	-0.0016 (0.0020)	-0.0013 (0.0021)	-0.00074 (0.0021)	0.0020 (0.0019)	0.000052 (0.0022)	-0.0039 (0.0021)	-0.00056 (0.0015)	-0.0050 (0.018)	0.024 (0.015)	0.0037 (0.0024)
Yearly income×Income data exists	-0.0000043 (0.0000043)	0.0000027 (0.0000045)	0.0000016 (0.0000042)	-0.00000019 (0.0000046)	-0.0000014 (0.0000043)	-0.00000049 (0.0000045)	0.0000070 (0.0000045)	-0.0000060 (0.0000029)	0.000077 (0.000038)	-0.0000069 (0.000030)	0.0000071 (0.0000056)
Income data exists	-0.0021 (0.0035)	-0.0059 (0.0034)	0.0080 (0.0035)	-0.0055 (0.0035)	-0.0025 (0.0033)	0.0046 (0.0036)	-0.0034 (0.0036)	0.0032 (0.0026)	-0.040 (0.030)	0.0088 (0.025)	-0.0022 (0.0041)
Any specialist care visit	0.0015 (0.0023)	-0.00052 (0.0022)	-0.0010 (0.0023)	-0.00057 (0.0023)	0.0016 (0.0021)	-0.00067 (0.0024)	0.0032 (0.0023)	-0.0017 (0.0017)	0.028 (0.020)	-0.0040 (0.016)	0.0014 (0.0027)
Any hospitalization	0.0020 (0.0028)	-0.0041 (0.0027)	0.0020 (0.0028)	-0.0029 (0.0028)	0.00089 (0.0027)	-0.0016 (0.0029)	-0.00062 (0.0029)	0.0030 (0.0021)	0.040 (0.025)	0.025 (0.020)	0.00081 (0.0033)
Patient paid copay	0.0039 (0.0029)	0.00015 (0.0028)	-0.0041 (0.0029)	0.00044 (0.0029)	0.0037 (0.0028)	-0.0030 (0.0031)	0.00081 (0.0030)	-0.0017 (0.0022)	0.044 (0.025)	0.035 (0.021)	0.0023 (0.0034)
Date-time shift FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	218,976	218,976	218,976	218,976	218,976	212,720	218,976	212,777	218,976	218,976	166,467
Joint p-value	0.14	0.38	0.42	0.43	0.09	0.96	0.59	0.41	0.30	0.39	0.37

Note: This table shows balance tests of doctor characteristics (columns) regressed against patient characteristics (rows), conditional on date-time shift fixed effects (FE). Date-time shift FE are a set of indicators for early morning, morning, afternoon, and evening for each specific date in the sample. The regressions are performed in the main sample of 218,977 consultations, but some patient or doctor characteristics may be missing, reducing the observation count. The variable "Income data exists" is included to account for individuals with no income data, such as children. Any specialist visit and any hospitalization indicate whether the patient had any specialist care visit or hospitalization during the three years prior to their online consultation. Standard errors are robust.

Table A6: Doctor-patient balance tests without fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	General Practitioner	Other specialty	Medical Doctor only	In specialist training	Completed specialist training	Lives in Stockholm Region	Speaks non-EU language	Top medical school graduation	Years of experience	Total consultations in all data (1000s)	Female doctor
Number of comorbidities	-0.0011 (0.0012)	0.0036 (0.0011)	-0.0025 (0.0012)	0.00030 (0.0012)	0.0022 (0.0011)	-0.0049 (0.0012)	0.0000044 (0.0012)	-0.00031 (0.00088)	-0.017 (0.010)	-0.037 (0.0087)	-0.0042 (0.0014)
Female patient	0.0013 (0.0021)	-0.0014 (0.0021)	0.000020 (0.0022)	-0.00091 (0.0022)	0.00089 (0.0020)	-0.0027 (0.0023)	0.0023 (0.0022)	-0.000034 (0.0016)	-0.033 (0.019)	-0.048 (0.016)	0.00041 (0.0026)
Patient age	0.00018 (0.000093)	-0.00033 (0.000089)	0.00015 (0.000093)	-0.00010 (0.000094)	-0.000052 (0.000088)	0.00060 (0.000099)	-0.00022 (0.000096)	0.00023 (0.000071)	0.00077 (0.00080)	0.0020 (0.00070)	0.00026 (0.00011)
1 st generation immigrant	-0.0039 (0.0033)	0.0017 (0.0032)	0.0022 (0.0034)	-0.00082 (0.0034)	-0.0014 (0.0031)	-0.00023 (0.0036)	0.00082 (0.0034)	0.00035 (0.0025)	-0.0071 (0.029)	-0.025 (0.025)	0.0053 (0.0040)
2 nd generation immigrant	-0.0061 (0.0042)	0.0057 (0.0041)	0.00040 (0.0042)	-0.0046 (0.0042)	0.0042 (0.0040)	-0.0061 (0.0045)	0.00015 (0.0043)	0.0027 (0.0032)	0.016 (0.036)	-0.012 (0.032)	0.0066 (0.0050)
Lives in Stockholm Region	0.0020 (0.0021)	-0.0022 (0.0020)	0.00021 (0.0021)	-0.0012 (0.0021)	0.00099 (0.0020)	0.0083 (0.0022)	-0.0031 (0.0022)	0.0024 (0.0016)	0.017 (0.018)	0.086 (0.016)	0.0094 (0.0025)
Yearly income×Income data exists	-0.0000040 (0.0000042)	0.0000061 (0.0000043)	-0.0000021 (0.0000043)	0.000000049 (0.0000046)	0.0000020 (0.0000042)	-0.0000032 (0.0000048)	0.000010 (0.0000048)	-0.000010 (0.0000031)	0.000069 (0.000038)	-0.000043 (0.000032)	-0.0000023 (0.0000056)
Income data exists	0.0012 (0.0035)	-0.031 (0.0033)	0.030 (0.0035)	-0.017 (0.0035)	-0.014 (0.0033)	0.044 (0.0037)	-0.026 (0.0036)	0.018 (0.0026)	0.043 (0.029)	0.10 (0.026)	0.018 (0.0042)
Any specialist care visit	-0.0021 (0.0023)	0.0018 (0.0022)	0.00037 (0.0023)	0.0043 (0.0023)	-0.0047 (0.0022)	0.0023 (0.0025)	0.0066 (0.0024)	-0.0035 (0.0017)	0.051 (0.020)	0.098 (0.017)	0.0058 (0.0027)
Any hospitalization	0.0017 (0.0029)	-0.0049 (0.0028)	0.0031 (0.0029)	-0.0042 (0.0029)	0.0011 (0.0027)	0.0028 (0.0031)	0.00040 (0.0030)	0.0030 (0.0022)	0.058 (0.025)	0.060 (0.022)	0.0053 (0.0034)
Patient paid copay	0.0018 (0.0028)	0.020 (0.0027)	-0.021 (0.0029)	0.0033 (0.0029)	0.018 (0.0027)	-0.026 (0.0030)	0.031 (0.0029)	-0.0025 (0.0022)	0.010 (0.024)	0.0064 (0.021)	-0.0013 (0.0034)
Observations	218,976	218,976	218,976	218,976	218,976	212,720	218,976	212,777	218,976	218,976	166,467
Joint p-value	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: This table shows balance tests of doctor characteristics (columns) regressed against patient characteristics (rows). The regressions are performed in the main sample of 218,977 consultations, but some patient or doctor characteristics may be missing, reducing the observation count. The variable "Income data exists" is included to account for individuals with no income data, such as children. Any specialist visit and any hospitalization indicate whether the patient had any specialist care visit or hospitalization during the three years prior to their online consultation. Standard errors are robust.

Table A7: Statistics on Emergency Room visits in sample

Panel A: Event chain and diagnosis

	Share	Observations
Share Emergency Room visits followed by a hospitalization (within two days)	0.14	12,102
Hospitalizations preceded by an Emergency Room (within two days)	0.55	4,437
Emergency Room visits within three days of online visit	0.52	12,138
Emergency Room visits with the same two-character diagnosis code as in the online visit	0.19	12,138
Emergency Room visits with the same diagnosis letter group as in the online visit	0.29	12,138

Panel B: 6 most common diagnosis categories of Emergency Room (captures 22%)

	Share	Observations
Abdominal and pelvic pain	0.099	1,160
Pain in throat and chest	0.031	368
Other soft tissue disorders, not elsewhere classified	0.025	294
Headache	0.023	273
Medical observation and evaluation for suspected diseases and conditions, ruled out	0.023	268
Dorsalgia (back-pain)	0.018	213

Panel C: NYU classification categories of Emergency Room

	Share	Observations
Unclassified	0.46	11,678
Failure I: Emergent condition, required care within 12 hours but could be treated in primary care.	0.17	11,678
Failure II: Non-emergent condition, the patient did not require immediate care within 12 hours.	0.15	11,678
Unclear: Injury	0.091	11,678
Not failure: Emergent condition, required Emergency Room care and was not preventable via ambulatory care.	0.075	11,678
Unclear: Psychiatric	0.031	11,678
Failure I: Emergent condition, required Emergency Room care but was preventable with timely ambulatory care.	0.029	11,678

Note: This table presents contextual information about the 12,138 Emergency Room visits observed in the analysis sample. Panel A describes the event chains associated with an Emergency Room visit: how they relate to hospitalizations and how closely the diagnosis relates to that of the online visit. Note that the observation number is different in the first two rows of Panel A. In the first row, this is due to truncating the sample by two days to observe a future hospitalization, and in the second row, it is because the estimates condition on having a hospitalization, of which there are 4,437 in the sample. Panel B displays the six most common three-character diagnosis codes occurring among the 12,138 Emergency Room visits. Panel C categorizes Emergency Room visits according to the Billings algorithm (NYU 2023). Just as in US data, not all diagnosis codes are classified by the algorithm, as indicated by the "Unclassified" category. In addition, my data use the Swedish ICD-10 code system, in which there are slight differences to the ICD-10-CM codes used by the Billings algorithm; these differences are reflected in the lower observation count. An emergent condition is one where the patient requires prompt medical care, either in primary care or at an emergency department. Emergent conditions requiring Emergency Room treatment are divided into those that could have been and those that could not have been prevented in primary care. For instance, flare-ups of asthma or diabetes require Emergency Room care but could have been prevented if timely ambulatory care had been provided during the episode of the illness. In contrast, trauma or appendicitis also requires Emergency Room care but could not have been prevented by ambulatory care treatment.

Table A8: Correlation between patient risk variables

	Counter-Guideline risk	Primary Care risk	Emergency Room risk
Counter-Guideline risk	1		
Primary Care risk	-0.11	1	
Emergency Room risk	-0.017	0.19	1
Mean of risk variable	0.13	9.39	0.73

Note: This table shows the correlation between unstandardized patient risk variables. For a description of these variables, see Section II.D.

Table A9: Autocovariance and autocorrelation vectors

	Counter-Guideline	Primary Care	Emergency Room
Lag 1	0.00018 (0.000032) [0.22] N = 4,980	0.00065 (0.00016) [0.11] N = 4,980	0.000079 (0.000023) [0.048] N = 4,980
Lag 2	0.00020 (0.000027) [0.24] N = 4,523	0.00029 (0.00016) [0.047] N = 4,523	0.00012 (0.000025) [0.072] N = 4,523
Lag 3	0.00016 (0.000030) [0.19] N = 4,177	0.00045 (0.00016) [0.070] N = 4,177	0.00010 (0.000025) [0.061] N = 4,177
Lag 4	0.00015 (0.000028) [0.17] N = 3,882	0.00034 (0.00012) [0.053] N = 3,882	0.000058 (0.000028) [0.034] N = 3,882
Lag 5	0.00013 (0.000026) [0.16] N = 3,621	0.00040 (0.00011) [0.063] N = 3,621	0.00011 (0.000029) [0.066] N = 3,621
Lag 6	0.00012 (0.000027) [0.15] N = 3,365	0.00016 (0.00010) [0.025] N = 3,365	0.000086 (0.000027) [0.052] N = 3,365
Lag 7	0.00012 (0.000023) [0.14] N = 3,156	0.00022 (0.00011) [0.033] N = 3,156	0.00014 (0.000028) [0.078] N = 3,156
Lag 8	0.000097 (0.000024) [0.12] N = 2,953	0.00029 (0.00010) [0.045] N = 2,953	0.000072 (0.000033) [0.041] N = 2,953
Lag 9	0.00013 (0.000030) [0.16] N = 2,780	0.00034 (0.000097) [0.050] N = 2,780	0.000074 (0.000033) [0.042] N = 2,780
Lag 10	0.000088 (0.000028) [0.11] N = 2,612	0.00013 (0.00010) [0.019] N = 2,612	0.000088 (0.000032) [0.050] N = 2,612
Lag 11	0.00012 (0.000030) [0.15] N = 2,461	0.00028 (0.00012) [0.0044] N = 2,461	0.000051 (0.000030) [0.028] N = 2,461
Lag 12	0.00011 (0.000039) [0.14] N = 2,295	0.00016 (0.00013) [0.023] N = 2,295	0.00013 (0.000034) [0.072] N = 2,295
Lag 13	0.00012 (0.000035) [0.14] N = 2,146	0.00036 (0.00012) [0.054] N = 2,146	0.000036 (0.000031) [0.020] N = 2,146
Lag 14	0.00013 (0.000036) [0.15] N = 2,024	0.00025 (0.00012) [0.037] N = 2,024	0.000063 (0.000032) [0.036] N = 2,024
Lag 15	0.000095 (0.000041) [0.11] N = 1,892	0.00015 (0.00014) [0.022] N = 1,892	0.000063 (0.000035) [0.035] N = 1,892
Lower bound of doctor skill s.d.	0.013	0.025	0.0089

Note: This table presents the estimated autocovariances between doctors' average residualized patient outcomes across two-week periods. It also shows the standard errors clustered at the doctor level in parentheses, the autocorrelation in brackets, and the number of doctors $\times$ time-period pairs observed with the specified number of lags below. Finally, the table shows an estimate of the lower bound of doctors' standard deviation of skill based on the individual doctor skill estimation; this is the square root of the first autocovariance, as in Chetty, Friedman, and Rockoff (2014).

Table A10: Outcome projection on patient characteristics

	Dependent variable ($\times 100$)		
	Counter-Guideline	Primary Care	Emergency Room
Share of antibiotics among past prescriptions (std)	0.22 (0.03)		
Number of past Primary Care prescriptions (std)		10.12 (0.36)	
Number of past Emergency Room visits (std)			2.85 (0.13)
Number of comorbidities	0.04 (0.03)	1.58 (0.13)	0.98 (0.07)
Female patient	-0.07 (0.06)	4.46 (0.19)	-0.92 (0.10)
Patient age	0.008 (0.002)	0.214 (0.007)	0.022 (0.003)
Constant	1.56 (0.07)	15.10 (0.30)	4.97 (0.12)
Outcome mean	1.78	25.07	5.54
Observations	218,977	218,977	218,977

Note: Patient-outcome-specific risk variables (in the first three rows) are standardized. For a description of the patient risk measures, see Section II.D. Standard errors are robust.

Table A11: Interaction effects between skill and patient risk with controls

	Dependent variable ($\times 100$)		
	Counter-Guideline	Primary Care	Emergency Room
Doctor skill (std) ($\hat{\beta}_1^k$)	-0.83 (0.05)	-0.81 (0.17)	-0.15 (0.07)
Patient risk (std) ($\hat{\beta}_2^k$)	0.21 (0.03)	10.13 (0.36)	2.97 (0.13)
Doctor skill $\times$ Patient risk ($\hat{\beta}_3^k$)	-0.09 (0.04)	-0.14 (0.34)	-0.45 (0.14)
Number of comorbidities	0.04 (0.03)	1.59 (0.13)	0.93 (0.07)
Female patient	-0.04 (0.06)	4.39 (0.19)	-0.84 (0.11)
Patient age	0.01 (0.00)	0.21 (0.01)	0.02 (0.00)
Constant	1.51 (0.07)	15.23 (0.31)	4.99 (0.13)
Date-time shift FE	✓	✓	✓
Observations	211,883	211,883	211,883
Outcome mean	1.76	25.0	5.52
Mean for low risk	1.70	22.2	4.81
Mean for high risk	2.22	67.3	13.6

Note: This table tests the robustness of the estimates shown in Table 6 to the inclusion of patient controls. Standard errors are bootstrapped with 1,000 iterations. For a description of the risk measures, see Section II.D.

Table A12: Explaining skill through doctor characteristics

	Panel A: Years of experience		
	Counter-Guideline	Primary Care	Emergency Room
Years of experience	-0.0277 (0.0047)	0.0000 (0.0039)	0.0009 (0.0043)
Observations	435	435	435
Outcome mean	-0.031	0.025	0.17
R^2	0.036	0.0027	0.010
Panel B: Medical school ranking			
Years of experience	-0.0279 (0.0047)	-0.0013 (0.0039)	0.0029 (0.0043)
Top medical school graduation	0.0506 (0.0481)	-0.0671 (0.0532)	0.2896 (0.0602)
Observations	426	426	426
Outcome mean	-0.035	0.014	0.17
R^2	0.040	0.0027	0.037
Panel C: Additional controls			
Years of experience	-0.0280 (0.0049)	-0.0025 (0.0041)	0.0039 (0.0044)
Top medical school graduation	0.0297 (0.0487)	-0.0640 (0.0544)	0.3035 (0.0617)
General Practitioner	0.0046 (0.0438)	0.0125 (0.0415)	-0.0259 (0.0384)
Total consultations in all data (1000s)	0.0347 (0.0065)	-0.0107 (0.0093)	-0.0681 (0.0352)
Speaks non-EU language	-0.0739 (0.0453)	0.0900 (0.0400)	-0.0869 (0.0432)
Urban residence	0.0661 (0.0411)	0.0042 (0.0356)	-0.0402 (0.0381)
Observations	424	424	424
Outcome mean	-0.037	0.012	0.17
R^2	0.051	0.0074	0.092
Panel D: Additional controls plus gender and Stockholm Region			
Years of experience	-0.0169 (0.0065)	0.0094 (0.0069)	0.0188 (0.0084)
Top medical school graduation	-0.1009 (0.0716)	-0.0141 (0.0803)	0.2121 (0.0826)
General Practitioner	-0.0279 (0.0551)	-0.0217 (0.0540)	0.0397 (0.0500)
Total consultations in all data (1000s)	0.0330 (0.0075)	-0.0149 (0.0104)	-0.0701 (0.0358)
Speaks non-EU language	-0.0889 (0.0556)	0.0870 (0.0512)	-0.1181 (0.0533)
Urban residence	0.0762 (0.0662)	-0.0006 (0.0599)	0.1053 (0.0664)
Lives in Stockholm Region	0.0012 (0.0680)	-0.0746 (0.0614)	-0.1313 (0.0676)
Female doctor	0.1553 (0.0497)	-0.0093 (0.0504)	-0.0117 (0.0488)
Observations	253	253	253
Outcome mean	-0.054	0.015	0.21
R^2	0.034	0.018	0.15
Panel E: Career level instead of years of experience with no controls			
In specialist training	-0.0324 (0.0412)	0.0478 (0.0478)	-0.1667 (0.0461)
Completed specialist training	-0.2433 (0.0471)	0.1379 (0.0457)	0.0155 (0.0430)
Observations	435	435	435
Outcome mean	-0.031	0.025	0.17
R^2	0.021	0.0088	0.020

Note: This table shows the relationship between observable doctor characteristics and estimated doctor skill in preventing the outcomes specified in column headers. Panel A includes years of experience, measured in reference to the doctor's graduation year, but not all doctors' graduation years are observed. To address missing values, I impute experience for the missings based on doctors' seniority (for details, see Section E.3). A variable used to flag the imputed values is also included in regressions where years of experience is used. Panel B adds as a regressor an indicator of graduation from Sweden's top medical school, which reduces the sample size as I only have that data on a sub-sample of doctors. Panel C adds additional controls, and Panel D adds doctors' gender and an indicator for whether they live in the wider Stockholm region, which further decreases the sample size. Panel E drops the sample-reducing controls and replaces the years of experience variable with seniority indicators. Variables are further described in Table B1. Standard errors are bootstrapped with 1,000 iterations.

Table A13: Doctor skill/outcome relationship

	Dependent variable ($\times 100$)		
	Counter-Guideline	Primary Care	Emergency Room
Doctor Counter-Guideline skill (std)	-0.835 (0.049)	0.693 (0.112)	0.096 (0.057)
Doctor Primary Care skill (std)	0.222 (0.062)	-0.844 (0.167)	-0.064 (0.062)
Doctor Emergency Room skill (std)	0.083 (0.052)	-0.044 (0.137)	-0.143 (0.064)
Date-time shift FE	✓	✓	✓
Outcome mean	1.76	25.03	5.52
Observations	211,883	211,883	211,883

Note: This table presents OLS estimates of how a 1 s.d. increase in each doctor skill affects the probability of the different healthcare outcomes. Each column specifies one of the three healthcare outcomes, and the rows specify (standardized) estimated doctor skills at preventing all three outcomes. Each element of the table is the coefficient of a separate regression. Doctor skill is standardized at the doctor $\times$ two-week time-period level. Standard errors are bootstrapped with 1,000 iterations.

Table A14: Mean distance (kilometers) between doctor and patient

	Counter-Guideline	Primary Care	Emergency Room
Panel A: Status quo conditional random allocation			
Rural residence $\times$ low risk	288	287	288
Rural residence $\times$ high risk	284	295	282
Urban residence $\times$ low risk	202	203	205
Urban residence $\times$ high risk	209	201	184
	Counter-Guideline	Primary Care	Emergency Room
Panel B: Positive assortative matching risk-skill allocation			
Rural residence $\times$ low risk	288	287	287
Rural residence $\times$ high risk	286	295	290
Urban residence $\times$ low risk	204	202	205
Urban residence $\times$ high risk	202	215	183
Observations	203,310	203,310	203,310

Note: This table shows the mean distance in kilometers from the patient's municipality centroid to the doctor's municipality centroid in the allocation that occurred in the online service, compared to a counterfactual positive assortative matching (PAM) on doctor skill and patient risk. Since ties may arise in the PAM, I randomize which of two equivalent patients in terms of risk are matched with which of two doctors; these matches are equivalent for the average reallocation effect but may differ in terms of distances. The point estimates presented in the table are means over 1,000 different PAMs. Column headers specify which outcome is used for PAM (higher-skilled doctors and higher-risk patients with respect to the outcome are matched together). The reallocations are the same as those reported in Figure 3. The patient types that specify the rows are defined by the patient's risk for each outcome and whether they live in an urban or rural municipality. A high-risk patient is defined as one with risk above 1 s.d. over the mean within two-week periods. The observation count is lower than in other tables due to some missing location data.

Table A15: Sample patients compared to national population

	Unweighted		National data weighted by	
	Sample	National	Municipality	Municipality ×gender×age
Urban residence	0.56	0.32	0.56	0.56
Female patient	0.61	0.50	0.50	0.61
Patient age	30.3	40.8	39.8	30.3
University education	0.55	0.38	0.43	0.50
Married	0.32	0.42	0.41	0.34
Immigrant background	0.17	0.24	0.27	0.30
Yearly income (1000s SEK)	323	311	333	318

Note: This table compares patients in the analysis sample to the full Swedish population (i.e., not to primary care patients specifically). The "Sample" column starts from the 218,977 observations and reports unweighted means for the non-missing demographic and socioeconomic variables. The missing pattern makes the sample vary between 146,145 for university education (defined only for patients aged above 23 years old) and 218,977. The "National" column is the mean for all inhabitants for whom data exists (i.e., also above 23 y.o. for education) in Sweden in 2017, as reported by Statistics Sweden in public data. The following two columns present the national data weighted based on the first set of characteristics in Column 1. The "Municipality" column uses the share of each of the municipalities in the sample (first column) as weights. The "Municipality×gender×age" uses the count of individuals in each municipality, gender, and age in the data as weights. "Immigrant background" is an indicator for a patient who is a first-generation immigrant or who is born in Sweden to two first-generation immigrants. "Yearly income" is measured in 2017 for patients in the analysis sample in this table, to match with the year when the national data are measured. See Table B1 for definitions of the remaining variables.

Table A16: Comparison of doctor characteristics to nationwide in-person public primary care doctors

	Sample: in specialist training	National: in specialist training	Sample: General Practitioner	National: General Practitioner
Doctor age <29	0.041	0.118	0.00	0.00
Doctor age 30-39	0.816	0.628	0.20	0.14
Doctor age 40-49	0.122	0.195	0.25	0.34
Doctor age 50-59	0.010	0.054	0.42	0.28
Doctor age > 60	0.010	0.005	0.13	0.24
Doctor female	0.371	0.591	0.52	0.58
N for gender (age)	105 (98)	2,030 (2,030)	61 (60)	2,650 (2,650)

Note: This table compares the age and gender profile of doctors in the analysis sample to the national population of doctors working in public primary care clinics. The comparison is reported separately for doctors in specialist training (ST) and for doctors who have completed their training in general practice. Doctors who only have MDs or are specialists in other fields are excluded, as they do not appear to be fully reported (e.g., due to missing data on medical locums and AT (residency) doctors) in the report by Olsson, Isacson, and Kolmodin (2022), from which the data on public primary care clinics are derived. In total, there are 435 doctors in the online care sample. Of these, 127 only have an MD, and 55 are non-GP specialists who completed their ST; these are excluded from the table. Furthermore, the relevant variables are only available for a subset of the 253 ST or GP doctors in the sample. Specifically, 66% of the doctors have gender data, and 62% have age data. Finally, the national data are only based on publicly provided primary care, while around half of primary care nationwide is provided by private providers.

Table A17: In-person primary care quality regressed on patients' deprivation index in Region Scania

	Positive experience	Satisfied with information	Met physician rather than other profession	Recommended treat- ment for diabetics
Deprivation index	-10.10 (2.11)	-5.72 (1.96)	0.02 (0.03)	0.40 (3.14)
Constant	89.06 (2.17)	79.72 (1.98)	0.44 (0.03)	63.38 (3.03)
Observations	134	134	148	127
R ²	0.14	0.06	0.01	0.00

Notes: The sample for this table is primary care centers in Region Scania, and the unit of observation is a primary care center with its registered patients. The regressor deprivation index is based on all patients registered with each primary care clinic, not only those who also use online primary care. This is possible because the deprivation index is from separate data; it is calculated by Statistics Sweden and used by administrative regions in order to risk-adjust capitation payments (see Appendix Section D.1 for a definition of the deprivation index). The outcome variables in Columns 1 and 2 are measures of satisfaction reported by patients at the clinic. Whether the patient met a physician rather than other profession such as a nurse is calculated from consultation-level Region Scania data. The recommended treatment for diabetics is lipid-lowering treatment, which is recommended by national guidelines to be added to standard diabetes treatment and chosen by Region Scania as a quality metric. Robust standard errors are in parentheses. Sources: Region Scania and Nationell Patientenkät (National Patient Survey or NPE), 2019.

Table A18: In-person primary care quality regressed on local inhabitant characteristics, nationwide

	(1)	(2)	(3)	(4)	(5)
Std(1 st generation immigrant)	-0.20 (0.03)		-0.17 (0.03)	-0.16 (0.03)	-0.16 (0.03)
Standardized mean income		0.15 (0.03)	0.09 (0.03)	0.09 (0.03)	0.09 (0.03)
Average patient age				0.01 (0.02)	0.01 (0.02)
Share of male				-0.03 (1.17)	-0.03 (1.18)
Std(1 st generation immigrant) $\times$ Standardized mean income					0.001 (0.02)
Region fixed effects	✓	✓	✓	✓	✓
Observations	1,270	1,277	1,248	1,248	1,248

Notes: The unit of observation is a four-digit postcode, which is matched to the location of primary care clinics. The outcome variable comes from *Nationell Patientenkät* (National Patient Survey or NPE, 2019). It is a standardized measure of the patients' average view of the primary care clinic's quality. Std(1st generation immigrant) measures the standardized share of foreign-born inhabitants in the municipality where the post code is located. "Standardized mean income" measures the standardized mean income in the municipality where the post code is located. "Avg. age" refers to the average age of inhabitants in the area, and "Share of male" measures the average share of male inhabitants. Region fixed effects are included. Robust standard errors are in parentheses.

Table A19: Duration and patient satisfaction for doctors skilled at preventing Counter-Guideline

	Patient time, mins	Admin time, mins	Patient gave doctor the top rating (missings set to zero)	Patient gave doctor the top rating (missings treated as missing)
High skill doctor	-0.066 (0.086)	-0.202 (0.205)	-0.028 (0.006)	-0.015 (0.006)
High risk patient	-0.137 (0.024)	-0.256 (0.060)	0.006 (0.004)	0.007 (0.004)
High skill doctor $\times$ High risk patient	-0.040 (0.076)	0.233 (0.202)	-0.012 (0.011)	-0.013 (0.012)
Date-time shift FE	✓	✓	✓	✓
Observations	207,127	207,123	211,883	154,963
Outcome mean	4.83	11.77	0.56	0.77

Note: This table shows whether doctors who are more skilled at preventing Counter-Guideline prescriptions take a longer time, or are differently evaluated by patients, than other doctors, particularly when seeing high-risk patients. High-risk patients (and high-skill doctors) are indicated by having a risk (skill) score at least 1 s.d. above the respective mean. As usual, the patient (doctor) risk (skill) distribution is defined across patients (doctors) within each two-week time period. The first two columns present information on the consultation duration partitioned into two parts: the time the doctor spent with the patient and the time the doctor spent on administrative tasks related to the consultation, such as writing prescriptions, referrals, or notes. The final two columns indicate how the patient scored the consultation. The outcome indicates whether the patient gave the consultation a perfect 5/5 score, but in the first of these columns, the missing values are set to 0, while in the second column, I condition on the patient having submitted a survey answer. Standard errors are bootstrapped with 1,000 iterations.

Table A20: Duration and patient satisfaction for doctors skilled at preventing Primary Care

	Patient time, mins	Admin time, mins	Patient gave doctor the top rating (missings set to zero)	Patient gave doctor the top rating (missings treated as missing)
High skill doctor	-0.112 (0.111)	-0.520 (0.204)	0.028 (0.012)	0.022 (0.011)
High risk patient	0.140 (0.037)	0.157 (0.083)	-0.050 (0.005)	-0.030 (0.006)
High skill doctor $\times$ High risk patient	0.064 (0.113)	0.347 (0.268)	0.018 (0.017)	0.025 (0.017)
Date-time shift FE	✓	✓	✓	✓
Observations	207,127	207,123	211,883	154,963
Outcome mean	4.83	11.77	0.56	0.77

Note: This table shows whether doctors who are more skilled at preventing Primary Care outcomes take a longer time, or are differently evaluated by patients, than other doctors, particularly when seeing high-risk patients. High-risk patients (and high-skill doctors) are indicated by having a risk (skill) score at least 1 s.d. above the respective mean. As usual, the patient (doctor) risk (skill) distribution is defined across patients (doctors) within each two-week time period. The first two columns present information on the consultation duration partitioned into two parts: the time the doctor spent with the patient and the time the doctor spent on administrative tasks related to the consultation, such as writing prescriptions, referrals, or notes. The final two columns indicate how the patient scored the consultation. The outcome indicates whether the patient gave the consultation a perfect 5/5 score, but in the first of these columns, the missing values are set to 0, while in the second column, I condition on the patient having submitted a survey answer. Standard errors are bootstrapped with 1,000 iterations.

Table A21: Duration and patient satisfaction for doctors skilled at preventing Emergency Room

	Patient time, mins	Admin time, mins	Patient gave doctor the top rating (missings set to zero)	Patient gave doctor the top rating (missings treated as missing)
High skill doctor	-0.067 (0.139)	-0.475 (0.258)	0.002 (0.009)	0.004 (0.008)
High risk patient	0.226 (0.033)	0.199 (0.071)	-0.024 (0.005)	0.008 (0.005)
High skill doctor $\times$ High risk patient	-0.186 (0.115)	0.016 (0.250)	-0.002 (0.017)	-0.006 (0.017)
Date-time shift FE	✓	✓	✓	✓
Observations	207,127	207,123	211,883	154,963
Outcome mean	4.83	11.77	0.56	0.77

Note: This table shows whether doctors who are more skilled at preventing Emergency Room visits take a longer time, or are differently evaluated by patients, than other doctors, particularly when seeing high-risk patients. High-risk patients (and high-skill doctors) are indicated by having a risk (skill) score at least 1 s.d. above the respective mean. As usual, the patient (doctor) risk (skill) distribution is defined across patients (doctors) within each two-week time period. The first two columns present information on the consultation duration partitioned into two parts: the time the doctor spent with the patient and the time the doctor spent on administrative tasks related to the consultation, such as writing prescriptions, referrals, or notes. The final two columns indicate how the patient scored the consultation. The outcome indicates whether the patient gave the consultation a perfect 5/5 score, but in the first of these columns, the missing values are set to 0, while in the second column, I condition on the patient having submitted a survey answer. Standard errors are bootstrapped with 1,000 iterations.

Table A22: Distribution of diagnoses for which antibiotics were prescribed against guidelines

Acute sinusitis	21.8%
Impetigo	17.7%
Cutaneous abscess	16.4%
Acute tonsillitis	14.5%
Acne	9.9%
Erysipelas	6.1%
Lyme disease (too broad spectrum)	5.9%
Acute upper respiratory infection	2.9%
Acute suppurative otitis media	2.1%
Diagnosis missing	1.2%
Acute cystitis (too broad spectrum)	0.7%
Acute bronchitis	0.6%
Pneumonia	0.1%
Number of Counter-Guideline	3,908

Note: This table shows the frequency of the diagnosis codes related to the prescription of antibiotics against guidelines in online care. For some conditions, it is against guidelines to prescribe antibiotics in online care since an in-person test is required. A Counter-Guideline prescription is defined as an antibiotic prescription for which the diagnosis code is in one of the diagnosis categories shown in this table and there was an antibiotic prescription, except for the two diseases where "too broad spectrum" is noted, for which the prescription is only Counter-Guideline if the antibiotic prescribed was of too broad spectrum compared to the recommended one.

Table A23: LASSO selection of variables for alternative risk prediction

Indicator(s) for:	Counter-Guideline	Primary Care	Emergency Room
Hospitalizations			
In 30 days before consultation	✓	✓	✓
In quarters 1–8 before consultation	✓: q3,4,5	✓: q1,5,7,8	✓: q5,6,8
In years 1–2 before consultation			✓: y2
Emergency Room Visits			
In 30 days before consultation	✓	✓	✓
In quarters 1–8 before consultation	✓: q6,7,8	✓: q2,4,5,7	✓: q1,2,3,4,5
In years 1–2 before consultation	✓: y2	✓: y2	✓: y1
Specialist Visits			
In 30 days before consultation		✓	✓
In quarters 1–8 before consultation	✓: q2,3,4,6,7	✓: q1,5,7	✓: q1,2,3,5,6,7,8
In years 1–2 before consultation	✓: y2	✓: y1	✓: y1,2
Antibiotic Prescriptions			
In 30 days before consultation	✓	✓	✓
In quarters 1–8 before consultation	✓: q1,2,4,5,6,8	✓: q1,2,3,4,5,6,7,8	✓: q3,4
Primary Care Prescriptions			
In 30 days before consultation		✓	✓
In quarters 1–8 before consultation	✓: q1,2,5	✓: q1,2,3,4,5,6,8	
Earlier Diagnosed Comorbidities (since 2013)			
Alcohol Abuse		✓	✓
Blood Loss Anemia		✓	
Cardiac Arrhythmias		✓	✓
Chronic Pulmonary Disease		✓	
Congestive Heart Failure	✓	✓	✓
Deficiency Anemia	✓		✓
Depression		✓	✓
Diabetes, Complicated			
Diabetes, Uncomplicated		✓	
Drug Abuse		✓	
Fluid and Electrolyte Disorders	✓	✓	
Hypertension, Complicated			✓
Hypertension, Uncomplicated	✓	✓	
Hypothyroidism		✓	
Lymphoma		✓	
Metastatic Cancer		✓	✓
Obesity	✓	✓	
Other Neurological Disorders		✓	✓
Paralysis	✓	✓	
Peptic Ulcer Disease Excluding Bleeding		✓	
Psychoses	✓		
Pulmonary Circulation Disorders		✓	
Rheumatoid Arthritis/Collagen Vascular	✓	✓	✓
Solid Tumor Without Metastasis	✓		✓
Valvular Disease			
Weight Loss		✓	
Demographic Variables			
<i>Urbanicity</i>			
Mixed municipalities	✓	✓	
Rural municipalities			
Large city municipalities	✓	✓	
Municipality missing			
<i>Other demographics and socioeconomics</i>			
Low-income working adult			
2 nd generation immigrant	✓	✓	
1 st generation immigrant	✓	✓	✓
Female patient		✓	✓

Continued on next page

Table A23 – continued from previous page

Indicator(s) for:	Counter-Guideline	Primary Care	Emergency Room
<i>Age Category</i>			
Child (5-12 years)		✓	
Teen (13-19 years)	✓		
Adult (20-39 years)	✓		
Middle age (40-59 years)		✓	
Senior (60+ years)		✓	✓
Patients' Self-Reported Symptoms			
Abdominal Pain			
Acne	✓		
Allergy	✓		
Asthma			
Cardiac Arrhythmia			
Chicken Pox	✓		
Cold and Flu	✓		
Cold Sores	✓		
Constipation			
Contraceptives	✓		
Cough	✓		
Diarrhea or Vomiting			
Erectile Dysfunction			
Eye Infection	✓		
Fever			
General Health			
Hair Loss			
Headache			
Hypothyroidism			
Indigestion or Heartburn			
Nail Problem	✓		
Nicotine Addiction			
Pmdd			
Poisonous Animal Bite	✓		
Postponement of Menstruation			
Prescription Renewal	✓		
Psychology Other			
Quit Smoking			
Revisit			
Simple Caretype 1			
Sinusitis	✓		
Skin Rash	✓		
Sleep Disorder			
Sore Throat	✓		
Spot Checker			
Stage Fright			
Stomach Gases	✓		
Urinary Tract Infection			
Missing	✓		

Note: The checkmarks indicate which variables the logit LASSO model selected to predict the outcomes specified in the column headers. Only the Counter-Guideline prediction model was allowed to select from patients' self-reported symptom indicators. The first quarter and first-year indicators exclude the 30 days just before the consultation, so the indicators for the 30 days before and the first quarter before cover disjoint sets of dates. The past comorbidities are diagnosed in outpatient or inpatient care between 2013 and the patient's first consultation with the provider. "Patients' self-reported symptoms" are the provider's internal labels for patients' self-reported symptoms. This variable is omitted for about 50% of the observations since the form for symptoms was only implemented some time into the sample period. See Table B1 for detailed variable descriptions of the other variables.

Table A24: Patient risk model comparison

Panel A: Counter-Guideline risk

	Lagged outcome	Health and demographics	Health, demographics, and symptoms
RMSE	0.30	0.13	0.13
Corr(Counter-Guideline risk, Counter-Guideline)	0.015	0.015	0.11
Share Counter-Guideline in risk Q_1		0.017	0.0032
Share Counter-Guideline in risk Q_2		0.016	0.016
Share Counter-Guideline in risk Q_3		0.018	0.019
Share Counter-Guideline in risk Q_4		0.021	0.033

Panel B: Primary Care prescription risk

	Lagged outcome	Health and demographics	Health, demographics, and symptoms
RMSE	0.49	0.41	0.41
Corr(Primary Care prescription risk, Primary Care)	0.27	0.34	0.35
Share Primary Care prescriptions in risk Q_1	0.13	0.13	0.12
Share Primary Care prescriptions in risk Q_2	0.19	0.16	0.16
Share Primary Care prescriptions in risk Q_3	0.25	0.24	0.24
Share Primary Care prescriptions in risk Q_4	0.46	0.47	0.48

Panel C: Emergency Room risk

	Lagged outcome	Health and demographics	Health, demographics, and symptoms
RMSE	0.23	0.23	0.23
Corr(Emergency Room risk, Emergency Room)	0.13	0.16	0.17
Share Emergency Room in risk Q_1		0.036	0.029
Share Emergency Room in risk Q_2		0.041	0.042
Share Emergency Room in risk Q_3		0.050	0.051
Share Emergency Room in risk Q_4		0.096	0.10

Note: This table compares three models for patient risk prediction. For a description of the standard patient risk measures based on lagged outcomes and used in Column 1, see Section II.D. The health and demographics model described in Column 2 is a logit prediction model constructed using LASSO, with details in Appendix Table A23. The model contains an additive selection (by LASSO) of past medical history and demographic variables. Past medical history variables contain indicators for the patients' history in the two years before the online consultation of Emergency Room visits, hospitalizations, non-acute specialist care visits, primary care prescriptions, and antibiotic prescriptions. They also contain indicators for 31 different comorbidities. Demographic variables include the urbanicity of patients' home municipalities, a low-income indicator, immigrant background indicators, age category indicators, and a gender indicator. The health, demographics, and symptoms model in Column 3 is also a logit model constructed using LASSO but contains an additive selection of indicators for the patient's self-reported symptom prior to their online consultation (see details in Appendix Table A23). The "Share in risk quartile" variable is omitted if there is insufficient variation in the risk variable to split the dataset evenly four-ways.

Table A25: Interaction effects between skill and LASSO predicted risk deciles

	Dependent variable ($\times 100$)		
	Counter-Guideline	Primary Care	Emergency Room
Doctor skill (std)	-0.07 (0.05)	-1.13 (0.33)	-0.02 (0.12)
2 nd risk decile	0.10 (0.05)	0.77 (0.33)	0.79 (0.15)
3 rd risk decile	0.64 (0.08)	1.64 (0.31)	
4 th risk decile	1.23 (0.11)	3.13 (0.34)	0.96 (0.17)
5 th risk decile	1.23 (0.11)	5.53 (0.36)	0.89 (0.18)
6 th risk decile	1.24 (0.12)	8.57 (0.37)	1.39 (0.19)
7 th risk decile	1.65 (0.13)	12.17 (0.39)	1.88 (0.19)
8 th risk decile	1.47 (0.12)	19.02 (0.40)	2.60 (0.20)
9 th risk decile	2.18 (0.12)	29.11 (0.41)	3.83 (0.21)
10 th risk decile	4.79 (0.16)	47.32 (0.41)	10.34 (0.26)
2 nd risk decile $\times$ Skill	-0.01 (0.07)	0.34 (0.40)	-0.17 (0.17)
3 rd risk decile $\times$ Skill	-0.38 (0.11)	0.37 (0.40)	
4 th risk decile $\times$ Skill	-0.86 (0.14)	0.31 (0.41)	0.13 (0.19)
5 th risk decile $\times$ Skill	-0.65 (0.12)	0.29 (0.41)	0.22 (0.20)
6 th risk decile $\times$ Skill	-0.56 (0.13)	0.40 (0.43)	0.24 (0.22)
7 th risk decile $\times$ Skill	-0.93 (0.15)	0.40 (0.44)	-0.09 (0.22)
8 th risk decile $\times$ Skill	-0.91 (0.15)	0.58 (0.48)	-0.18 (0.22)
9 th risk decile $\times$ Skill	-1.14 (0.16)	0.31 (0.49)	-0.49 (0.24)
10 th risk decile $\times$ Skill	-2.07 (0.22)	0.67 (0.49)	-0.66 (0.31)
Constant	0.34 (0.06)	12.35 (0.23)	3.19 (0.11)
Date-time shift FE	✓	✓	✓
Observations	211,883	211,883	211,883
Outcome mean	1.76	25.0	5.52
Mean for low risk	1.41	19.4	4.75
Mean for high risk	4.60	52.7	15.7

Note: This table shows estimates of a alternative match effects to those described in Equation 12 and estimated in Table 6. In this table, I use more sophisticated measures of patient risk estimated using a logistic regression model of patient demographics and past healthcare use selected through a LASSO procedure (see Section F.1.2 for details). Second, I allow the LASSO-predicted risk variable to flexibly interact with doctor skill by using indicators for each decile of the risk distribution. In the bottom rows' descriptives of sample means, high-risk patients are those with a LASSO-predicted risk at least 1 s.d. above the mean; the rest are considered low-risk. Doctor skill is standardized at the doctor $\times$ two-week time-period level. Standard errors are bootstrapped with 1,000 iterations. See Section II.C for definitions of the outcome variables.

Table A26: Robustness not using value added, with smaller sample

	Dependent variable ($\times 100$)		
	Counter-Guideline	Primary Care	Emergency Room
2nd Risk Tercile	0.59 (0.35)	7.51 (0.57)	1.94 (0.24)
3rd Risk Tercile	0.74 (0.17)	26.54 (0.55)	7.43 (0.43)
2nd Skill Quartile	-0.58 (0.21)	-0.59 (0.57)	0.02 (0.21)
3rd Skill Quartile	-0.80 (0.24)	-1.96 (0.54)	0.09 (0.19)
4th Skill Quartile	-1.11 (0.23)	-1.60 (0.57)	0.20 (0.17)
2nd Risk Tercile $\times$ 2nd Skill Quartile	-0.34 (0.39)	-0.56 (0.76)	0.33 (0.39)
2nd Risk Tercile $\times$ 3rd Skill Quartile	-0.09 (0.41)	-0.30 (0.76)	-0.13 (0.37)
2nd Risk Tercile $\times$ 4th Skill Quartile	-0.35 (0.39)	0.12 (0.73)	-0.34 (0.36)
3rd Risk Tercile $\times$ 2nd Skill Quartile	-0.63 (0.24)	0.42 (0.73)	-0.31 (0.56)
3rd Risk Tercile $\times$ 3rd Skill Quartile	-0.32 (0.22)	0.68 (0.71)	-1.63 (0.59)
3rd Risk Tercile $\times$ 4th Skill Quartile	-0.22 (0.21)	0.15 (0.77)	-1.37 (0.57)
Constant	2.11 (0.17)	15.33 (0.44)	3.78 (0.12)
Date-time shift FE	✓	✓	✓
Observations	155,777	155,777	155,777
Outcome mean	1.70	24.9	5.31

Note: As a robustness check to the main methodology, this table presents interaction effects using alternative skill estimates performed in line with the methodology of the appendix of Adhvaryu et al. 2024. The respective skill is defined in this table as the negative of the mean outcome observed in each doctor's first third of the two-week time-periods in the sample. The presented estimates are performed on the remaining two-thirds of the data. For the regressions, doctors and patients are categorized based on their skill and risk into quartiles and terciles, respectively (see Section F.2 for details). Standard errors are clustered on doctors.

A2 Appendix figures

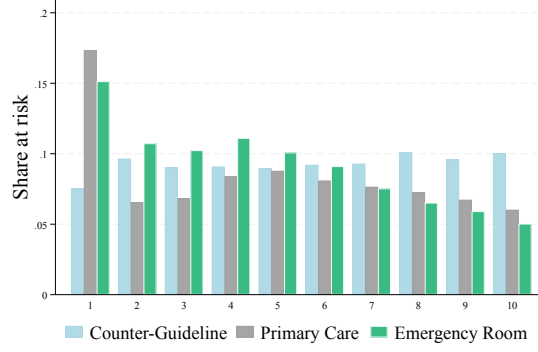
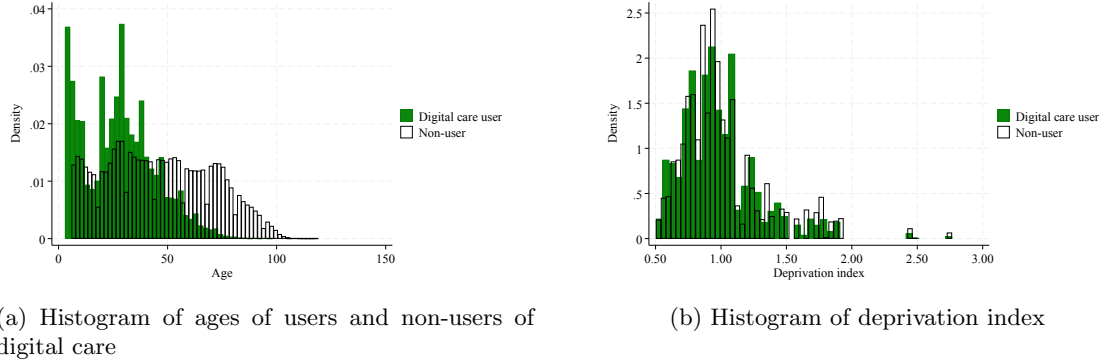


Figure A1: Proportion at risk over income deciles

Note: This figure shows the share predicted to be at risk for the different adverse health outcomes across patient income deciles (for those with non-missing income, i.e., almost all working-age adults). Patients are defined to be “at risk” if their respective outcome risk score is at least one standard deviation above the mean. See the main text for a description of the variables, particularly Section II.D. $N = 158,980$.



(a) Histogram of ages of users and non-users of digital care

(b) Histogram of deprivation index

Figure A2: Age and socioeconomic representativeness

Note: This figure shows selection into online primary care in Region Scania by age and socioeconomic status. Scania is the region where I have data on the full population of in-person primary care users and thus can study selection into the service. Because I do not have individual socioeconomic data on in-person care users who do not use online care, I use the deprivation index of the in-person primary care clinic where they are registered for all patients in panel (b). See Appendix Section D.1 for a definition of the deprivation index (obtained from Region Skåne (2017)).

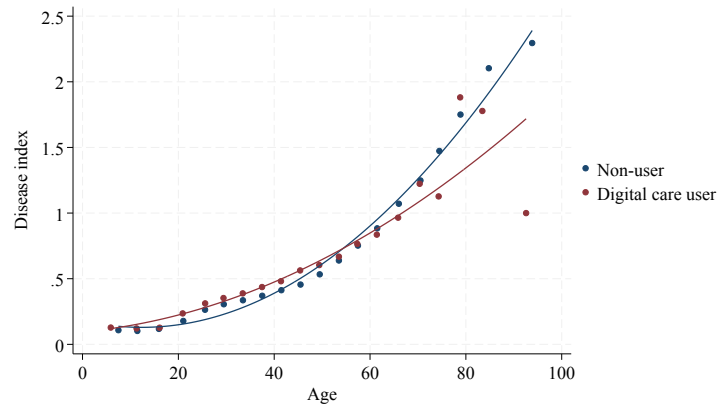
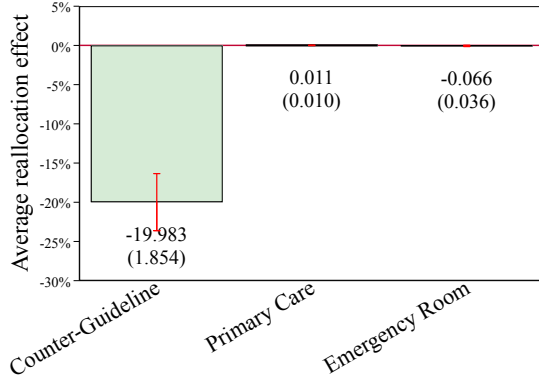
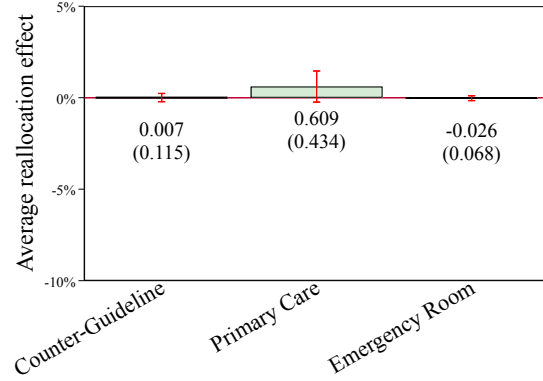


Figure A3: Comorbidity representativeness

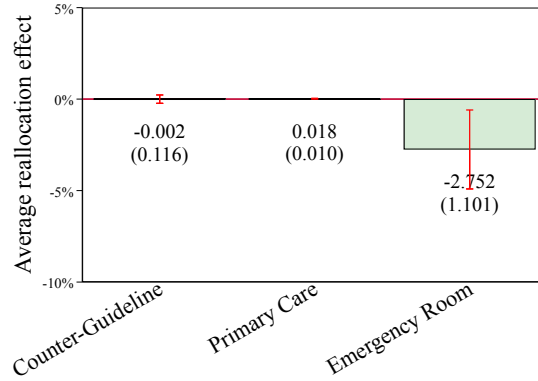
Note: This figure shows selection into online primary care in Scania by number of important comorbidities over age. It plots the sum of the comorbidity index (counting how many of the 31 comorbidities patients have had in primary care) over patient age, for users versus non-users of online care, using data from 2013 to 2016 in Region Scania, prior to the introduction of the online service, where I have primary care data on the full population.



(a) Minimizing Counter-Guideline outcomes



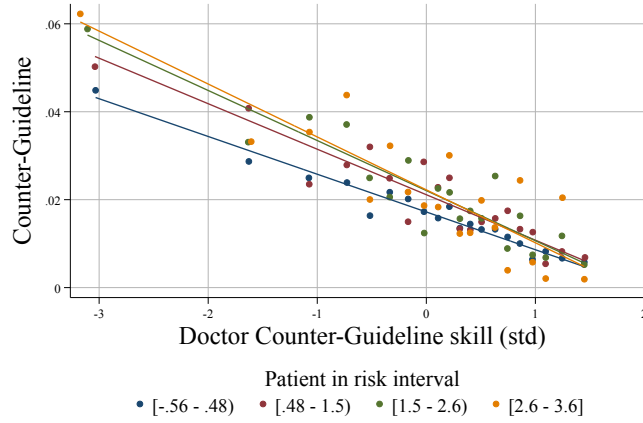
(b) Minimizing Primary Care outcomes



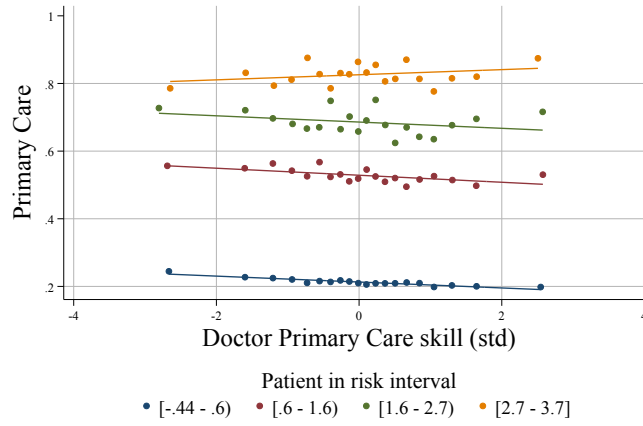
(c) Minimizing Emergency Room outcomes

Figure A4: Average reallocation effects from matching on skill- and LASSO-predicted risks

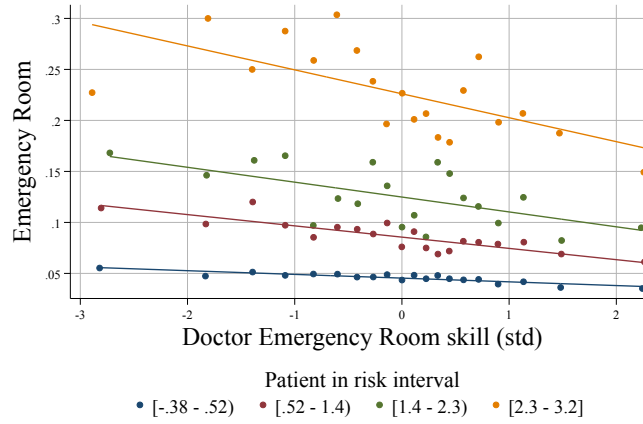
Note: This figure displays alternative average reallocation effects, with 95% confidence intervals and standard errors in parentheses. It modifies the approach taken in Figure 3 by leveraging a more complex patient risk measure created using a logit model selected through LASSO. See Section F.1.2 for details on the LASSO-predicted risks. Moreover, the average match function is estimated semi-parametrically using risk deciles. In each panel, the positive assortative matching of patients and doctors is implemented within each two-week period. The confidence intervals are bootstrapped with 1,000 iterations. $N = 211,883$.



(a) Counter-Guideline



(b) Primary Care



(c) Emergency Room

Figure A5: Binscatter of doctor skill and patient risk interaction effects

Note: This figure nonparametrically explores the average match function, Equation 6, between patient risk and doctor skill using binscatter plots. The patient risks are first winsorized at the 99th percentile to mitigate the influence of extreme values. The winsorized risk measures are then divided into four equal-length intervals based on their empirical ranges. For each outcome, a separate binscatter plot is produced where the x-axis represents the standardized doctor skill and the y-axis represents the corresponding adverse health outcome of patients. For each patient risk interval, a line is fitted on the binscatter estimates to highlight how the impact of doctor skill on patient outcomes varies across different patient risk levels. $N = 211,883$.

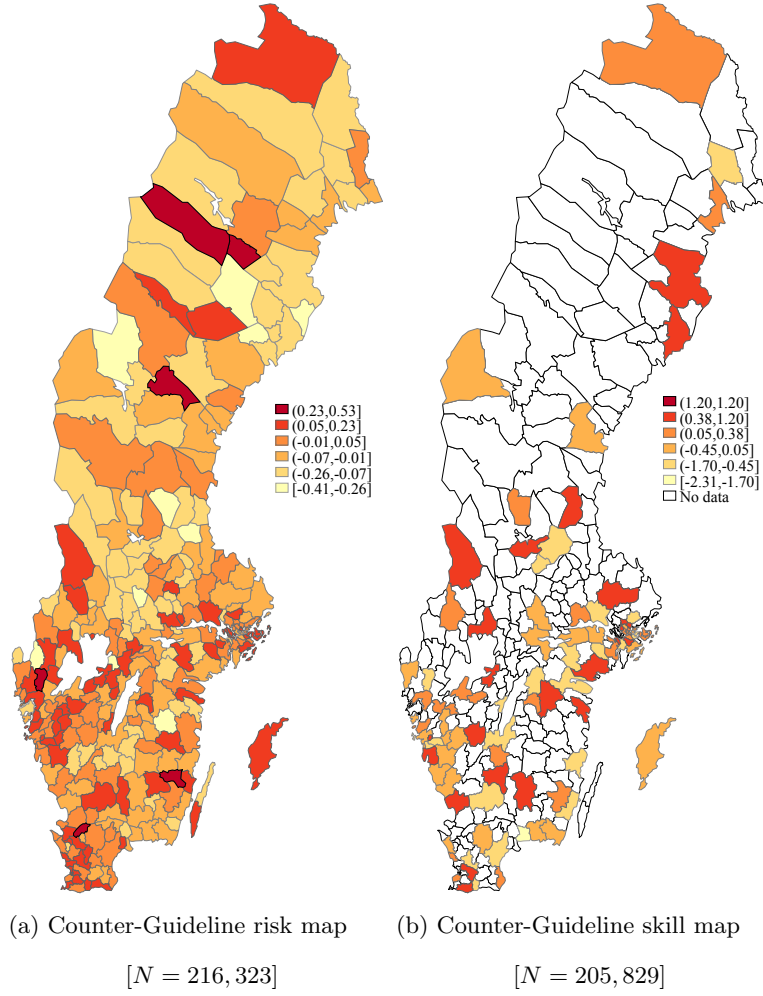


Figure A6: Counter-Guideline risk and skill distribution over Swedish municipalities

Note: This figure shows the distribution of patient Counter-Guideline risk and doctor Counter-Guideline skill over Sweden's 290 municipalities. Location data are missing for some doctors and patients, yielding a lower observation count than in the main analysis. The observation count reported in panel (b) is the number patient observations used to calculate the value added for doctors included in the figure, i.e., who have location data.

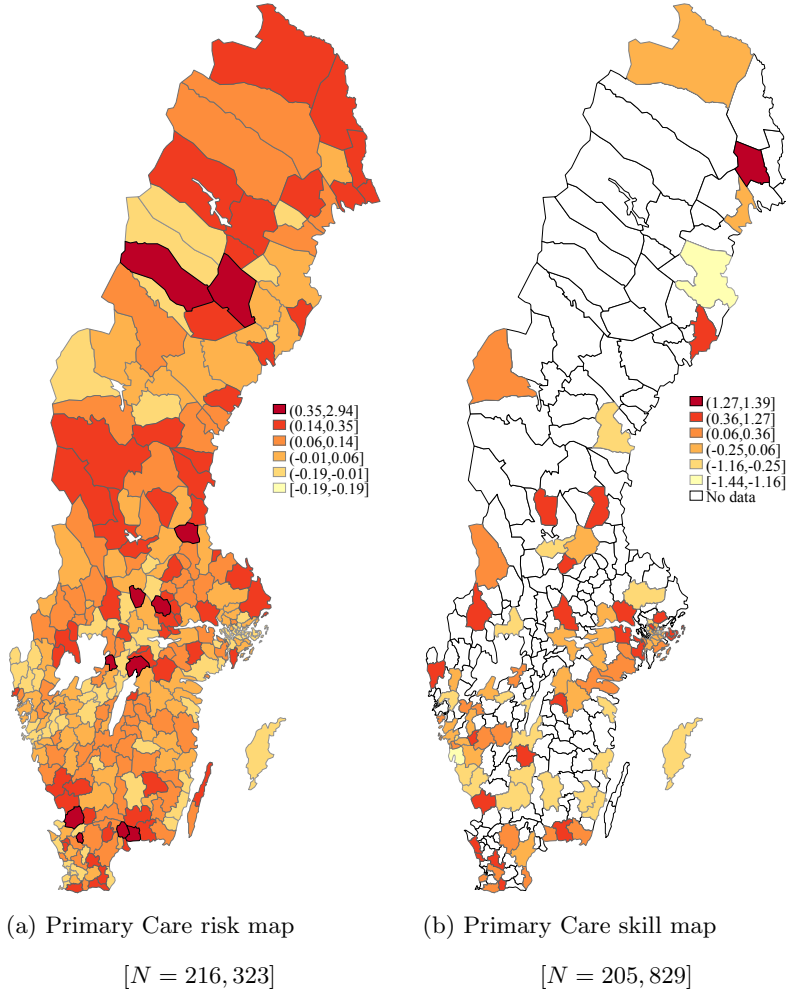


Figure A7: Primary Care risk and skill distribution over Swedish municipalities

Note: This figure shows the distribution of patient Primary Care risk and doctor Primary Care skill over Sweden's 290 municipalities. Location data are missing for some doctors and patients, yielding a lower observation count than in the main analysis. The observation count reported in panel (b) is the number of patient observations used to calculate the value added for doctors included in the figure, i.e., who have location data.

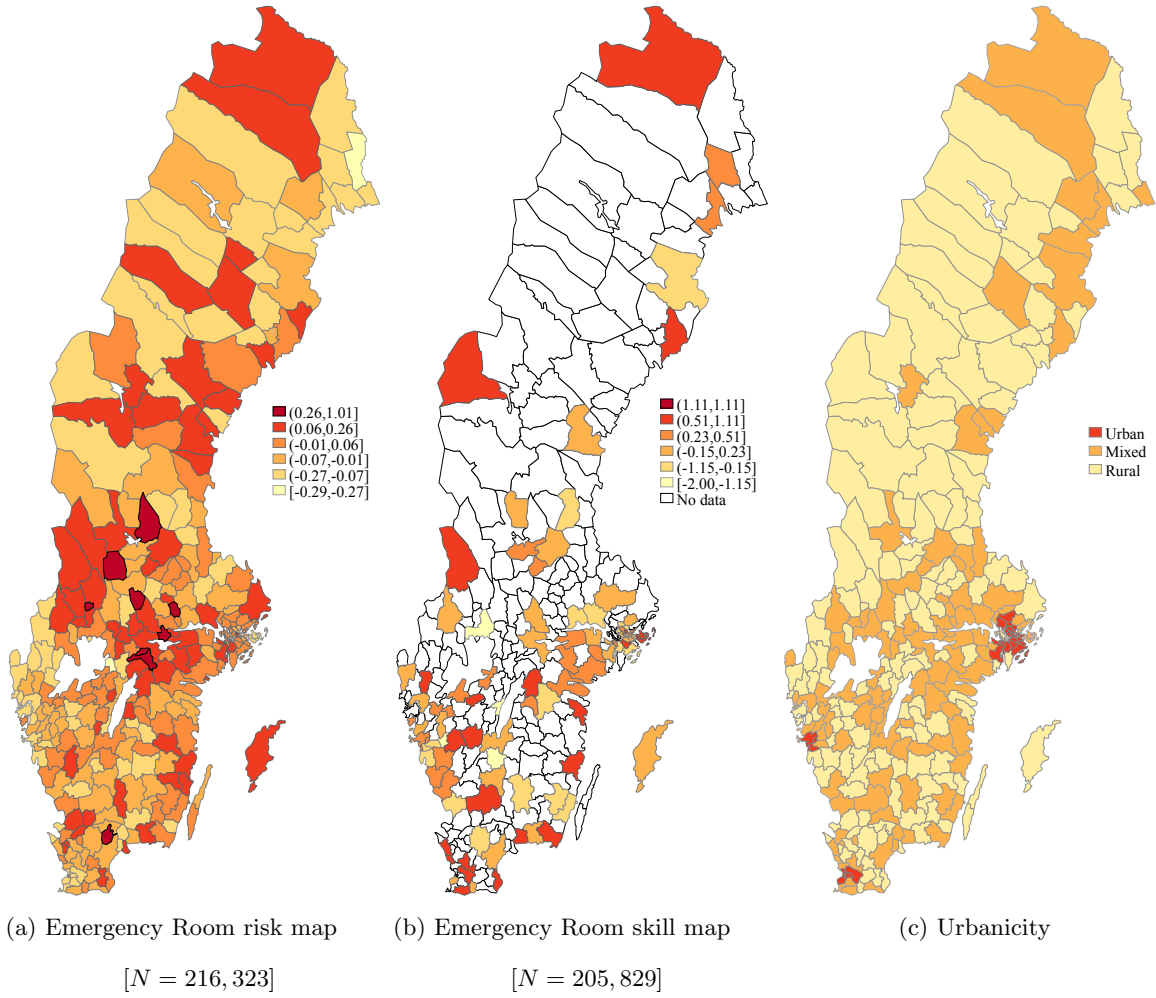


Figure A8: Emergency Room risk and skill distribution over Swedish municipalities and urbanicity

Note: This figure shows the distribution of patient Emergency Room risk and doctor Emergency Room skill over Sweden's 290 municipalities. Location data are missing for some doctors and patients, yielding a lower observation count than in the main analysis. The observation count reported in panel (b) is the number of patient observations used to calculate the value added for doctors included in the figure, i.e., who have location data. Panel (c) displays the urbanicity of municipalities in Sweden for comparison.

B Appendix: Variables

Table B1: Variable Descriptions

Variables	Descriptions
Outcome Variables	
Counter-Guideline	An indicator that the doctor prescribed antibiotics against national guidelines. See Appendix Section E.2 for details. Information about the drug code is registered when the patient collects the drug. An observation window of 30 days is used for drug collection.
Primary Care	An indicator that the patient picked up a prescription prescribed by a different primary care organization in the 30 days following the online doctor's consultation (possibly on the same day as the consultation). See Appendix Section E.2 for details.
Emergency Room	An indicator that the patient had an outpatient Emergency Room visit within 30 days of the doctor's consultation (possibly on the same day as the consultation). The Emergency Room definition includes hospital associated emergency rooms, specialty emergency clinics (e.g., psychiatric clinics), as well as urgent care centers. About 70% of the Emergency Room visits in the sample are hospital associated. See Section E.2 for details.
Patient risk proxies	
Counter-Guideline risk	The share of antibiotics among prescriptions in the three years before the consultation. This variable is standardized when used in analysis (not in the summary statistics table) over the patients.
Primary Care risk	The number of primary care prescriptions in the three years before the consultation. This variable is standardized when used in analysis (not in the summary statistics table) over the patients.
Emergency Room risk	The number of Emergency Room visits observed between 1 January 2016 and the time of the consultation (as the data only allows the Emergency Room visits to be distinguished from the other non-primary care outpatient visits from 2016 onward). Since most of the online consultations in the sample take place during 2017 and 2018, 96% of patients have Emergency Room data for at least one year before their online consultation; 85% are observed for at least a year and a half, and 66% are observed for at least two years. Calendar-date-by-time-shift fixed effects eliminate concerns about differential observation windows, since patients observed on the same calendar day have identical prior windows for Emergency Room visits. This variable is standardized when used in analysis, but not in the summary statistics table.
High-risk	A high-risk patient in each of the above risk categories has a risk variable which is at least one standard deviation above the mean. The remaining patients are defined as low-risk. This variable is defined separately for each risk variable (Counter-Guideline, Primary Care and Emergency Room risks).
Estimated skills	
Counter-Guideline skill	The estimated skill of a doctor with respect to preventing Counter-Guideline prescriptions. This variable is standardized at the doctor $\times$ two-week time-period level.
Primary Care skill	The estimated skill of a doctor with respect to leaving patients satisfied with their consultations, so that they do not get treatment from another primary care provider within 30 days of their online consultation. This variable is standardized at the doctor $\times$ two-week time-period level.
Emergency Room skill	The estimated skill of a doctor with respect to preventing patients from utilizing Emergency Room care in 30 days after their online consultation. This variable is standardized at the doctor $\times$ two-week time-period level.
High-skill	A high-skilled doctor in each of the above risk categories has an estimated skill which is at least one standard deviation above the mean.

Table B1: Variable Descriptions (Continued)

Variables	Descriptions
Fixed effects	
Date-time shift fixed effects (FE)	A factor variable for the date by year (i.e., exact calendar date, e.g., 3 March 2018) interacted with the four time "shifts" of the day. The shifts are: Night: 12am-6.59am, Morning: 7am-11.59pm, Afternoon 12pm-4.59pm, Evening: 5pm-11.59pm. See Section E.2 for details.
Patient demographics and socioeconomic characteristics	
Female patient	An indicator for the gender of the patient. Female patients are denoted by 1, male patients by 0.
Patient age	This variable denotes the age of the patient in 2017.
1 st generation immigrant	An indicator for whether the patient is a first-generation immigrant, that is, the patient was born outside of Sweden.
2 nd generation immigrant	An indicator for whether the patient is a second-generation immigrant born in Sweden, whose two parents were both born outside of Sweden.
Immigrant background	An indicator for a patient who is either a first- or second-generation immigrant.
Married	Indicator capturing the patient's civil status. Married is denoted by 1, divorced by 0. Patients younger than 18 are considered ineligible for marriage or divorce and are therefore set to missing.
Yearly income	The total (yearly) income from employment and self-employment (not capital income), derived from tax registries, for adult patients aged above 21. This variable is recorded in 2015 and expressed in thousands SEK.
Income data exists	An indicator for whether there is an observation for a patient's income. Takes the value one when a patient is old and young enough to work (age 20 to 67 in 2015) and has non-missing income data. Patients who are ineligible and missing income data are coded as 0.
Lives in Stockholm Region	An indicator for whether the patient lives in the wider Stockholm region. This is the most populous of Sweden's 21 regions and is home to the capital, Stockholm.
University education	This variable is based on patients' educational attainment in 2018 and is defined only for individuals aged 23 or older to ensure completion of education.
Patient paid copay	An indicator for whether the patient paid the national health insurance copayment for the visit; nonpayment reflects exemption due to age (child or elderly) or having reached the annual copayment ceiling.
Patient gave doctor the top rating (missings set to 0)	An indicator for whether the patient gave a score of 5 on a 1-5 scale to their consultation with the doctor, asked in a voluntary online post-consultation survey. Missings are coded as 0.
Patient gave doctor the top rating (missings treated as missing)	An indicator for whether the patient gave a score of 5 on a 1-5 scale to their consultation with the doctor, asked in a voluntary online post-consultation survey. Missings treated as missing.
Patient health	
Number of comorbidities	This variable denotes the number of clinically important comorbidities diagnosed in specialist (inpatient and outpatient) care during the three years prior to the patient's online visit. Importance is defined according to the Elixhauser comorbidity index (see Appendix Section E.2).
Any specialist care visit	An indicator for whether the patient had any specialist care visits within three years prior to their online consultation.
Any hospitalization	An indicator that the patient had a hospitalization within three years prior to their online consultation.

Table B1: Variable Descriptions (Continued)

Variables	Descriptions
Doctor characteristics	
Years of experience	Years of experience is defined as the difference between 2018 and the physician's graduation year, capturing total work experience. Graduation year is missing for approximately 50 percent of physicians due to incomplete coverage in the national education registries. For these cases, experience is imputed using physician seniority from provider records (MD only, specialty trainee, or specialist), which is fully observed. We apply conservative lower-bound assumptions: physicians with an MD only are assigned two years of experience (reflecting post-MD general service and primary care work required before being hired at the online service), specialty trainees are assigned five years of experience (two years of general training plus three years, corresponding to the midpoint of specialty training), and specialists are assigned eight years of experience (two years of pre-specialty training and six years of specialty training).
Top medical school graduation	An indicator for whether a doctor obtained their degree from the top medical school in Sweden (Karolinska Institutet). There are 7 medical schools in Sweden.
General Practitioner	An indicator for doctors who are either specialists in general practice (family medicine) or undergoing specialist training in general practice.
Other specialty	An indicator for all other doctors, except General Practitioners, who are specialists or undergoing specialist training.
In specialist training	An indicator for whether a doctor is training to become a specialist.
Completed specialist training	An indicator for whether a doctor has completed their specialist training.
Medical Doctor only	An indicator for doctors who have not yet started their specialist training.
Total consultations in all data (1000s)	A count (in thousands) of a clinician's total online consultations within the provider, including both consultations in the analysis sample and those with patients outside the sample. This variable captures the clinician's overall workload intensity within the provider.
Number of consultations per doctor	A count of the online consultations a clinician has had in the analysis sample.
Number of consultations per doctor $\times$ two-week pair	A count of a clinician's online consultations in the analysis sample within each two-week period.
Speaks non-EU language	An indicator for whether a physician speaks a language not among the official national languages of the first 15 EU member states (which include, e.g., English, Swedish, French, German, Spanish). In the sample, the most common non-EU languages are Arabic and Persian.
Female doctor	An indicator for the gender of the doctor. Female doctors are denoted by 1, male doctors by 0.
Other variables used only in the Supplemental Appendix	
Urban residence	An indicator for whether the patient or the doctor resides in an urban municipality as defined by the Swedish Agency for Economic and Regional Growth (Tillväxtverket).
Urbanicity	A categorical variable classifying the patient's or the doctor's municipality of residence into three levels: Urban, mixed and rural.
Deprivation index	The deprivation index (also called CNI) is a primary care clinic-level measure calculated by Statistics Sweden and used by Region Scania. It is a weighted average of seven indicators: (1) individuals aged 65 or older living alone; (2) immigrants born outside the EU; (3) unemployed individuals aged 16–64; (4) single parents with children under age 18; (5) individuals aged one or older who have moved into the area; (6) low educational attainment among individuals aged 25–64; and (7) children under age five.
Nationwide in-person primary care clinic quality score (patient reported)	This variable measures the patients' views of the primary care clinic's quality, taking on values between 0 and 100. Obtained from Nationell Patientenkät (Nationell Patientenkät (Socialstyrelsen) 2016; Nationell Patientenkät (Socialstyrelsen) 2018)

C Representativeness

1 Representativeness of the patient sample

Comparison of online patients to in-person primary care patients To compare sorting patterns into the online primary care service with sorting patterns into in-person primary care, I study one Swedish region where I have access to the universe of in-person primary care data.¹ This is Scania (Region Skåne), the southernmost administrative region in Sweden, containing both rural areas and the country’s third-largest city. Fourteen percent of digital care users come from this region, and 13% of Swedish inhabitants live there. The region is similar to the country overall in terms of education and income.²

Figure A2, panel (a) shows that digital care users in this region are considerably younger than in-person care users. However, conditional on age, digital and in-person patients under age 60 – who comprise the majority of digital care users – exhibit similar prior disease burden (Figure A3). Prior disease is measured by the number of comorbidities in previous primary care using the Elixhauser index, a commonly used measure for summarizing disease burden (Elixhauser et al. 1998).

Beyond health status, digital and in-person care users are also similar along socioeconomic dimensions. Using a clinic-level deprivation index, I find that the distributions of socioeconomic status are comparable across digital and in-person patients (Appendix Figure A2, panel b).³

Comparison of online care patients to Swedish inhabitants I also compare the digital care sample to the overall Swedish population on measures where this is possible (i.e., where public national data exist for variables available in the study sample). Table A15 shows a comparison between patients who used digital care during 2016–2018 and are included in my sample and residents of Sweden as a whole. Note that the municipality and national data refer to the population overall, not to primary care patients. A larger proportion (56%) of my sample comes from metropolitan municipalities compared to the population nationwide. Therefore, the second column contains

¹Primary care data are not collected by the national authority (the NBHW), which otherwise contributes the rest of the in-person healthcare data for this study. To obtain access to in-person primary care data nationwide, separate applications must be made to each of the 21 regions. I received data on the full universe of in-person primary care patients in Region Scania, representing nearly the entire population of the region. I do not have data on individual socioeconomic variables, such as income, or other healthcare than primary care for patients who do not use online care in this region, but I can link them to their in-person primary care clinic, for which some aggregate socioeconomic data exist.

²In Scania, 34.2% of people aged 25–64 have at least three years of postsecondary education, compared to 33.7% nationally. Scania’s median annual income per adult is ~326,000 SEK, compared with the Swedish median of ~342,000 SEK (Ekonomifakta 2023).

³This index is used by administrative regions and defined by Statistics Sweden. It is a weighted average of the following variables: born outside the EU; unemployed (ages 16–64); single parent with child under 18; low education (ages 25–64); over age 65 and living alone; over age one and moved into the area; and children below five. I do not have access to the underlying individual-level variables for non-online care users. Anell et al. (2024) show that low education, unemployment, and immigrant status are the most common reasons for assigning individuals to the deprived group. They also show that individuals assigned to the high-deprivation group also have considerably lower labor income, suggesting the index succeeds in distinguishing socioeconomic status. The data are provided by Region Skåne (Region Skåne 2017).

a weighted average of municipality-level data, weighted by the proportion of patients from each municipality in my sample.

There is a higher proportion of women than men in the online patient sample, while women make up half of the country’s total population. Women also tend to use more healthcare and primary care than men in general, so this pattern is not specific to online care but reflects a broader pattern among primary care users. In Region Skåne in 2017, 59% of all medical appointments were with female patients—a similar proportion as in my sample. Patients in the sample are also younger than the average resident. The service is particularly popular among parents (it is the child’s age that is registered) and less popular among the elderly.

Young women have higher rates of university education than men, which may partly explain why the percentage of individuals with university education is higher in my sample than in the country overall. Another contributing factor is that the percentage of university-educated individuals is higher in the municipalities where the patients come from.

The percentage of married individuals is lower in the online patient sample than in both the country and the municipalities, which can be partly explained by the younger sample, even if children are excluded from the “married” variable. There is also a lower proportion of individuals with an immigrant background in the sample compared to both the national and municipal populations. Average income in the sample is lower than in the municipalities (partly due to the younger age structure) but higher than in the country overall (which can partly be explained by the overrepresentation of metropolitan regions).

2 Comparison of sample doctors to nationwide in-person primary care

Doctors working in the online service are recruited across a range of experience levels, under the conditions that they (1) hold certification as a medical doctor (MD, *legitimerad läkare*) from the Swedish NBHW, which requires completion of 18–21 months of internship/early residency (Allmäntjänstgöring, “AT”) following the 5.5-year-long medical school program, and (2) have completed at least six months of their AT in a Swedish GP clinic/primary care center or have at least six months of experience working at a Swedish GP clinic after completing AT. After AT, Swedish doctors complete at least 5.5 years of specialty training (ST), during which they also continue to work as qualified medical doctors. Doctors who specialize in primary care/general practice/family medicine during their ST become GP specialists upon completion.

I compare the seniority distribution among doctors in the analysis sample to data from a report by Olsson, Isacson, and Kolmodin (2022) on the profiles of physicians in all *publicly provided* in-person primary care in Sweden.⁴ There appear to be more doctors with only an MD in my sample compared to national in-person primary care, but this could be due to under-reporting, as medical locums are not included in the report about in-person public primary care. Another difference is that there are more specialists in fields other than primary care/GP in my sample,

⁴Sixty percent of in-person primary care is publicly provided, and there are no data available on doctor profiles for privately provided, in-person primary care.

often in specialties such as acute medicine and internal medicine.

Table A16 compares seniority (during specialty training (ST) versus after ST), as well as age and gender patterns, between doctors in my sample and doctors from Olsson, Isacson, and Kolmodin (2022). There are fewer female physicians in my sample than in national public primary care, both among GP specialists and among doctors still in ST. This may reflect that women are more likely to work for public employers and select the GP specialty for its family-compatible schedule, whereas online work—which often involves nonstandard hours—may be more attractive to men. Among doctors in specialty training, age patterns are quite similar between my sample and national public primary care, though my sample has fewer very young doctors (under 29), slightly fewer doctors aged 40–49, and more doctors in the most common ST age group of 30–39. Among GP specialists, ages are also fairly similar, though there are more doctors aged 30–39 and 50–59, and fewer over age 60, in my sample compared to the national distribution.

D Additional results and background information

1 In-person primary care: Socioeconomic inequality

Primary care data in Sweden are not collected nationally but are instead held by the 21 administrative regions and by individual private or public providers, which makes it difficult to study primary care quality nationwide. In addition, while doctor identifiers are available in the online care data used in the main analysis, they are not available in nationwide in-person healthcare data. These data limitations are a key reason why nationwide online care substantially improves our ability to study doctor quality. Nevertheless, in this section, I present suggestive evidence on disparities in access to high-quality in-person primary care and assess whether there may be gains from trade.

Table A17 reports regressions using in-person primary care data from the administrative region Scania, where I observe the full population of primary care users rather than only online care users. The dependent variables are measures of primary care clinic quality. Outcomes in Columns 1 and 2 are drawn from the *Nationell Patientenkät* (Socialstyrelsen) (2018), while Columns 3 and 4 use publicly reported quality measures from Region Scania. On the right-hand side, I include a clinic-level measure of patient socioeconomic status – the deprivation index – constructed as an average of characteristics of patients registered with each clinic and described in Section C.1. The table shows that patients at clinics with higher deprivation index report a more negative experience with primary care (Column 1). Moreover, patients at clinics with higher deprivation are less satisfied with the information they receive in in-person primary care (Column 2). There is no significant relationship between deprivation and the share of patients who see a doctor rather than another health professional (e.g., a nurse) when visiting primary care (Column 3). Column 4 measures an aspect of objective quality of care – whether patients diagnosed with diabetes also receive lipid-lowering treatment – but shows no significant correlation with the deprivation index.

To assess whether the link between primary care quality and patient socioeconomic status holds nationwide, I use public clinic-level data from the National Patient Survey (*Nationell Patientenkät*

(Socialstyrelsen) 2016). Unlike the earlier analysis, the income and immigration measures used here are based on residential location rather than clinic registration. The advantage of this dataset is its national coverage – Table A18 includes most of Sweden by matching municipality-level and four-digit postcode-level data. The outcome is a patient-reported clinic quality score. Table A18 shows that clinic quality by this measure is lower in areas with lower incomes and in areas with higher shares of first-generation immigrants.

1.1 How patients choose in-person primary care clinics

In-person primary care is provided at primary care centers, financed through the universal public health insurance and operated either by private companies or by administrative regions. Ninety-nine percent of Swedish inhabitants live within 20 minutes of their nearest primary care clinic (Tillväxtverket 2011). Most patients are registered with one such clinic but usually not with an individual doctor. Patients also have the option to choose their clinic. However, research indicates that few patients actively choose their clinic and only 6% of individuals with low education select a center other than their assigned default, compared to 29% among those with higher education (Bendz 2011). The usual default is the closest clinic within the municipality.⁵

The lower likelihood of active choice among lower-education patients aligns with research showing that lower-income students are less responsive to quality when choosing schools (Bau 2022). In an experiment aimed at reducing information frictions and hassle costs in the Swedish primary care clinic choice, even when patients were prompted to make an active choice, they did not switch to clinics with higher quality, on average (Anell et al. 2023).

2 Doctors’ time use and patients’ satisfaction with doctors

Tables A19, A20 and A21 show regressions of the patient-facing meeting duration (first column), and the physicians’ related administration time (second column) on doctor high-skill and patient high-risk indicators and their interaction. The third and fourth columns of these tables regresses patient satisfaction on doctor high-skill and patient high-risk indicators and their interaction. Doctors with higher skills do not spend significantly more time with patients on average, or differentially more face time with high-risk patients, for any outcome. Thus, they do not achieve better outcomes by using more time resources. Moreover, as can be seen in the third and fourth columns of these tables, for Emergency Room and Counter-Guideline outcomes, there is no evidence that patients in general rate skilled doctors more highly or that high-risk patients give higher ratings to doctors better suited to them. This suggests that matching systems or suggestions/nudges are needed to improve allocation, as patients are generally unaware of doctor-specific skills for these outcomes. For Primary Care outcomes, however, there is evidence that patients recognize skill: patients rate doctors who are skilled at preventing these outcomes more highly.

⁵In some regions, such as Stockholm, patients can remain unregistered with any clinic if they do not make an active choice, while most regions a default clinic is assigned.

3 Informational requirements for reallocations

The informational requirements to carry out the reallocations consist of having access to patients’ past healthcare records. This relates to earlier research showing that electronic medical records can reduce deaths by making information more accessible (Miller and Tucker 2011). For predicting patient risks, data on past prescriptions and Emergency Room visits can be accessed in theory, as providers can have access to the national patient oversight tool, which collects electronic medical records.

The required data to measure doctor skills are doctors’ consultation histories and the outcomes of those visits. In the case of Counter-Guideline outcomes, the outcome data already exist within the medical provider’s system, as both diagnoses and prescription drugs are recorded and can be used to determine guideline adherence. For Emergency Room and Primary Care outcomes, one-month follow-up data would be needed for patients seen by each doctor whose performance is assessed. This could be achieved through integration with existing medical records. For instance, patients with subsequent Emergency Room visits could be flagged. Such feedback would be useful even in the absence of a reallocation objective. Currently, ownership of the means of prediction remains with the governmental agencies hosting patient data as well as with the providers that produce it.

E Data appendix

1 Definition of the analysis sample

I start from the universe of doctor consultations with the largest provider of online healthcare in Sweden (and Europe), covering the period from the start of the service in Sweden in 2016 to the end of 2018. There are 631,664 consultations in this original dataset (see Table A2) with potentially several observations per patient – on average, patients had 1.67 online consultations. I keep only the first visit for each patient, as using later visits could introduce concerns about endogeneity. Hence, each patient has only one observation in digital care in the analysis sample. After also excluding some test consultations prior to the official opening of the service in 2016, this leaves 376,270 consultations.

I match these data to official registry data from Statistics Sweden on socioeconomic and demographic variables and to data from the NBHW on visits from specialist, acute, and inpatient care, and on prescriptions from all care, across the Swedish healthcare system in 2013–2018. This time period covers three years before digital primary care was introduced and the entire digital care sample period. The in-person healthcare dataset contains many observations (consultations, hospitalizations, prescriptions, etc.) per patient. All patients from the online care data match with the official registry data, though some variables may be missing for certain patients.

I further restrict the sample to drop-in visits – those where patients cannot select a specific doctor but instead meet the first available doctor. These 309,777 visits comprise 82% of the first-visit sample, and this is the sample where time-conditional random assignment holds. The largest share

(approximately half) of the remaining 18% of visits involved patients booking a specific doctor, a feature mainly available during the early days of the service and which was almost completely phased out in early 2018. Most of the drop-in sample, where conditional random assignment holds, comes from 2018, when drop-in was the main pathway. In addition, comparing observable characteristics such as gender, age, immigration background, yearly income, and consultation satisfaction (measured by top ratings), I find no large differences between patients who booked a specific doctor and those who went through the conditional random assignment.

Moreover, I remove pediatricians since first-come, first-served randomization did not hold for them, as a preference was built in for pediatricians seeing small children. For the same reason – and to ensure a consistent amount of pre-digital care data for everyone (2013–2016), I remove small children.⁶ This leaves 233,470 patients and 499 doctors.

I also implement some restrictions to ensure that controls and outcomes can be included for all patients. First, I require that primary control variables (age, gender, number of comorbidities) be jointly defined for all observations. Second, I truncate the sample at the end of 2018 to allow a one-month post-consultation observation period for each patient, as the datasets measuring outcomes also end in the end of 2018.

When estimating doctor skills, I partition the data into two-week periods. To estimate value added, each doctor must be observed in at least two time periods. This leaves 218,977 observations, which I use to estimate value added. Since I only use past data to estimate value added, the first time period for each doctor cannot receive a value-added estimate, leaving 211,883 observations for subsequent analysis using the estimated value added. This is the main analysis sample.

Finally, for the estimation of the distribution of doctor skills, with results in Table 4, I restrict the data further by requiring that every doctor-by-time-period cell contains at least 30 consultations.⁷ This leaves 172,177 observations, when I have also applied the restriction that there should be at least two time periods per doctor again. For the value-added estimates or for the reallocation effects using the value-added estimates, restricting to high volume does not matter.⁸

⁶In practice, I remove all children born after 2012 to ensure that children in the sample are at least three years old at any point they are observed in digital care (which started in 2016). This fulfills both the condition that the remaining children do not have a pediatrician preference built into their assignment and allows for a consistent definition of patient types based on their pre-digital in-person healthcare use.

⁷Rose, Schellenberg, and Shem-Tov (2022) drop the parts of teachers' observations where the teacher has less than 15 students in that subject. In other research, a minimum threshold of 30 cases per individual is a common restriction (Frandsen, Lefgren, and Leslie 2023).

⁸Coefficients based on value added which underlie reallocation effects, and which usually use the main sample, are robust to restricting to high-volume time periods (as above), and also to high-volume doctors – see further Supplemental Appendix Tables A3 and A4. When estimating value added using only past time periods, there is never a value added assigned to the first time period, so 172,177 observations to estimate value added leaves 161,135 observations in regressions that use value-added estimates, such as Table A3 .

2 Important variables

Counter-Guideline outcomes

As explained in Section II.C, Counter-Guideline prescriptions are coded based on 16 national guidelines relating to online care from Strama, the Swedish strategic program against antibiotic resistance (Strama 2017; Strama 2019). These guidelines aim to limit the use of antibiotics when unnecessary or to encourage the use of narrow-spectrum antibiotics when they are considered effective. This aims to limit the use of broad-spectrum antibiotics, which contribute more to the negative consequences of antibiotic resistance. Thus, to follow the guidelines, doctors sometimes need to decline requests for antibiotics from patients who believe they need them. The non-adherence measured in the sample is quite low (1.8%) by international standards, as is common in Scandinavia. Ellegård, Kjellsson, and Mattisson (2022) show that online care providers in Sweden are more restrictive in prescribing antibiotics than in-person providers.

To define the variable, I combine information on prescriptions in the digital care data, conditional on the diagnosis (ICD) code, with data on drug codes from the NBHW’s prescription register, which records the Anatomical Therapeutic Chemical (ATC) code of the prescriptions once patients have filled them. The vast majority of Counter-Guideline antibiotics prescriptions are picked up from the pharmacy within only a few days after the online consultation: 96% are picked up in 7 days and 98% in 14 days. However, I measure prescriptions filled within 30 days to maintain consistency with the other outcomes.

Table A22 shows the distribution of diseases associated with Counter-Guideline outcomes in the data. For Lyme disease and acute cystitis, prescribing a narrow-spectrum antibiotic is guideline-adherent, and a prescription is classified as Counter-Guideline only if a broad-spectrum antibiotic was used. For the remaining diagnoses in the table, any antibiotic prescribed in online care is against the guidelines. To avoid measuring coding errors rather than treatment errors, I do not code missing ICD codes associated with antibiotics prescriptions as Counter-Guideline, unless the doctor explicitly sets the ICD code as “Diagnosis missing,” which accounts for a very small proportion of all Counter-Guideline prescriptions (around 1%).

Primary Care outcomes

The outcome variable for Primary Care outcomes measures whether patients received additional treatment in in-person primary care, based on prescription data from the NBHW. I have data on all prescriptions picked up by patients between January 2013 and December 2018, including the type of healthcare provider (e.g., primary care) that prescribed the drug. I use this to create an indicator of whether the patient received a prescription from another primary care provider (i.e., outside the online care company) in the month after the online doctor consultation. I use prescriptions as a measure of additional primary care treatment because data on primary care consultations are not collected nationwide, only data on primary care prescriptions.⁹

⁹There are both type I and type II errors with this variable if one aims to measure whether patients have a subsequent in-person primary care consultation. I assess the extent of these errors in the region where I have

Emergency Room outcomes

Emergency Room outcomes are measured within 30 days of the online primary care visit, as for the other outcomes. However, the majority (52%) of visits occur within three days of the online primary care consultation (Appendix Table A7). Twenty-nine percent share the same diagnosis group as the online visit. In healthcare in general, it is common for diagnosis groups to change over a patient’s care journey, for example, from a symptomatic group to a confirmed diagnosis. The most common diagnosis categories for Emergency Room visits are abdominal and pelvic pain, pain in throat and chest, and other soft tissue disorders.

The New York University Billings algorithm NYU (2023) divides Emergency Room visits into different types, in a similar way as the algorithm dividing hospitalizations into avoidable and non-avoidable hospitalizations. The Billings algorithm was developed by medical professionals who classified Emergency Room diagnosis codes into different categories. Unfortunately, around half of all diagnosis codes are left unclassified by the algorithm in US settings, and this is also true in this paper’s data (46%). Appendix Table A7 describes which groups the classified visits fall into, and this is also discussed in the main text of the paper, Section II.C. That section also discusses the different types of failure in primary care the Emergency Room visits indicate. There are two groups in the table indicating the first type of failure in primary care (Failure I), in which a condition was not properly treated in primary (also called ambulatory) care: “Emergent condition, required care within 12 hours but could be treated in primary care” and “Emergent condition, required Emergency Room care but was preventable with timely ambulatory care”. The sum of their probabilities is $0.17+0.029 = 0.199$. However, there are several categories that are unclear or unclassified ($46\%+9.1+3.1\%=58.2\%$), so among those clearly classified, which I call interpretable, (42.4%), the share of Failure I is 47% ($19.9\%/0.424$). The second failure is due to the patient seeking Emergency Room care for a non-emergent condition, which makes up 15% of the total and 35% of the interpretable Emergency Room visits. This should ideally be prevented by the patient being properly informed about the right management and level of care by the online primary care doctor. Together, the two failures make up over 80% of the interpretable Emergency Room visits.

Patient comorbidities

Throughout the paper, I often control for whether a patient has any important comorbidity, meaning they have at least one important diagnosis (out of a pre-defined list) recorded in prior healthcare visits in the national registries maintained by the NBHW. Important comorbidities are defined by 31 families of ICD-10 diagnosis codes. A patient is assigned a comorbidity if they were diagnosed with any of these ICD-10 codes during any inpatient, outpatient, or hospitalization encounter (not

complete primary care data. Type I errors (false positives) occur when patients collect prescriptions prescribed in primary care without having an in-person doctor visit that month.

Conversely, type II errors (false negatives) occur when patients have a primary care consultation without collecting a prescription. However, these errors seem to mostly offset each other in the sample of online care patients. In Scania, where I can observe all primary care visits, 9.1% of Scania patients in the analysis sample have a physical appointment in other primary care within 30 days of the online consultation without collecting a prescription, and a similar proportion, 9.8%, collect a prescription from in-person primary care without having a physical appointment.

primary care data as that is not collected nationwide).

The list of 31 comorbidities (ICD-10 code families) includes congestive heart failure; cardiac arrhythmias; valvular disease; pulmonary circulation disorders; peripheral vascular disorders; hypertension, uncomplicated; paralysis; other neurological disorders; chronic pulmonary disease; diabetes, uncomplicated; diabetes, complicated; hypothyroidism; renal failure; liver disease; peptic ulcer disease excluding bleeding; AIDS/HIV; lymphoma; metastatic cancer; solid tumor without metastasis; rheumatoid arthritis/collagen vascular; coagulopathy; obesity; weight loss; fluid and electrolyte disorders; blood loss anemia; deficiency anemia; alcohol abuse; drug abuse; psychoses; depression; and hypertension, complicated.

This list is taken from the Stata command *elixhauser* written by Stagg, Hilsden, and Quan (2015), though I adapted their code to fit the present setting. In particular, while the original *elixhauser* implementation uses four-character ICD-10 codes, the external patient records available here include only three-character ICD-10 codes. I therefore adapt the code families to include the three-character parent of any four-character code used by Stagg, Hilsden, and Quan (2015). For example, family Pulmonary Circulation Disorders were originally defined by Stagg, Hilsden, and Quan (2015) as codes “I26,” “I27,” “I280,” “I288,” and “I289”; in my implementation, the family is defined by “I26,” “I27,” and “I28.” The definition thus includes a broader set of diagnoses for some groups, but not for all. Finally, the variable I use measures the patient’s number of comorbidities, defined as the number of comorbidity families with which the patient’s earlier medical diagnoses are associated.

Demographics

The demographic and socioeconomic variables come from the Statistics Sweden dataset. The demographic variables, such as immigration background and municipality, are available for almost all patients. However, I can only measure most socioeconomic characteristics for adults, as the variables on income and education, for example, do not exist for minors. The socioeconomic variables from Statistics Sweden are measured at the same time for all individuals, irrespective of the year when they started using the online healthcare service. Table B1 describes the variables.

Control variables

There are four important control variables that I employ in most of the estimates in this paper. Three of these variables are in some tables indicated by a check mark in the row labeled “Controls”: the comorbidity score (explained above), patient age, and patient gender. I also often condition on a fourth variable, “Date-time FE” (fixed effects), which is a factor variable for the date-specific “shift” in which the online doctor consultation occurred. Each exact date (e.g. 1st of June 2018) is partitioned into four shifts: Night: 12am-6.59am, Morning: 7am-11.59pm, Afternoon 12pm-4.59pm, Evening: 5pm-11.59pm.

Doctor-patient distances

In Table A14, I provide information on distances between doctors and patients. Exact locations are not available; instead, I use the centroid of each individual’s municipality as their location. Municipality data are available for almost all patients and for 424 doctors (97.5%).

3 Data sources

All datasets are confidential and were obtained following approval from the Stockholm Regional Ethics Council (2018, no. 2108/2318-31) and the Swedish Ethics Authority (2019, no. 2019-06062). Additionally, Statistics Sweden and the other entities carried out their own confidentiality assessments before approving data sharing. Statistics Sweden matched with datasets, anonymized the personal identifiers, and then shared only an anonymized version of the data with the researcher through secure sharing. Much of the data are considered sensitive since they contain health-related information.

For the main analysis, I draw on data from a number of different sources. The starting point is consultation-level data from the largest digital healthcare provider in Sweden in 2016–2018. These data are proprietary and merged with (1) administrative individual patient and doctor-level panel data from Statistics Sweden’s Integrated Database for Labor Market Research (LISA), covering 2013–2018 (some only until 2017) for patients, and 2016–2022 for doctors; (2) healthcare panel data on all patient visits from 2013 to 2018; including inpatient and outpatient care data, provided by the NBHW; and (3) prescription data for all patients at the drug level (i.e., each prescribed and collected drug appears as an observation line from 2013 to 2018, again provided by the NBHW), in-person primary care data from one region, and doctor data. These are described below. A few additional datasets used for supplementary analysis are described in the readme file of the replication package.

Online primary care provider data. The online consultation-level data between 2016 and 2018 provide the backbone for the analysis. For each consultation, I have data on start time (hourly precision), duration (precision of seconds), doctor anonymous identifier and doctor characteristics, visit type (e.g., drop-in visit), patient fees, post-consultation satisfaction survey responses, and consultation outcomes. Two important visit outcomes is the patient’s diagnosis, recorded using ICD-10-SE codes with four to five characters of precision, and whether the patient received a prescription.

Demographics and socioeconomic data on patients. To complement the primary care data, I obtain demographic and socioeconomic micro-data on patients from Statistics Sweden, drawn from the Integrated Database for Labor Market Research (LISA). These panel data provide information on individual income, educational attainment, municipality of residence, immigration background, and marital status. The variables are provided at the patient level, with annual measurements. The income measurement is a summation for the year, the education attainment is measured at

the end of the year’s spring semester, immigration background is constant, and municipality and marital status are measured on either December 31 or January 1 of the following year.

In-person healthcare data. The NBHW provides in-person healthcare data, which contain data from the entire healthcare system – both from public and private providers – except primary care consultations. These data, covering 2013–2018, may be divided into inpatient and outpatient care data. Outpatient care includes Emergency Room visits and other non-primary care visits, such as specialist visits. However, Emergency Room visits can only be distinguished from other non-primary care outpatient visits from January 2016 onward.

These data contain one or several ICD-10 diagnostic codes for each consultation, with three characters of precision. They also include external causes codes for applicable cases, classifying events such as falls and bites, and provide visit dates (and discharge dates for inpatient visits). However, they do not contain anonymous clinician identifiers.

Prescription data. The prescription data are also provided by the NBHW. These data encompass all prescriptions that patients collected between 2013 and 2018, including those prescribed in primary care. I can tie prescriptions back to the online consultation, and for each drug picked up, there are data available on its Anatomical Therapeutic Chemical (ATC) code with full precision.

In-person primary care data. Since the national registries from the NBHW do not contain primary care consultation data, I applied separately for primary care data from the largest regions in Sweden and successfully received it in its entirety from one region, Region Skåne. I have data on in-person primary care (2013–2019) on the universe of primary care appointments (i.e., both for patients who used online primary care and for all other patients in the region), which allows me to study selection into the online service based on variables available in the Region Skåne data.¹⁰

Doctor data. Data on doctors come from two data providers: the online care firm and Statistics Sweden. The within-firm data include each doctor’s specialty, seniority, spoken languages, age, and gender. I obtain the remaining data—such as graduation year, the municipality where the medical degree was obtained, and the municipality of residence (which also defines the coarser category region)—from Statistics Sweden.

From the online provider, two important variables are doctors’ seniority and specialization. Doctors fall into four categories: (1) those who have only completed their MD (i.e., medical degree plus residency (AT)), (2) those in specialist training, (3) those who have completed their specialist training in general practice, and (4) those who have completed their specialist training in any specialty. I use these variables in e.g. balance checks and to provide comparisons to national-level data in Table A16. By counting the number of consultations each doctor conducted, I generate the

¹⁰I do not have complete socioeconomic and demographic data on patients who are not in the online care sample, and so I cannot study selection into online care on all measures. Instead, I use age, comorbidities, and the deprivation index, which is calculated by Statistics Sweden for the pool of patients at in-person primary care clinics in this region.

“Total consultations in all data (1,000s)” variable, reflecting the number of consultations (including those outside the analysis sample) during 2016–2018.

I use the Statistics Sweden data to define additional doctor characteristic variables. Using the municipality of the medical degree, I infer whether the degree was obtained from Karolinska Institutet (the top medical school in the country out of seven medical schools, and the only medical school in its municipality). I calculate each doctor’s years of experience as the difference between 2018 and their last graduation year. To ensure consistency, I flag doctors whose years of experience are inconsistent with their age (specifically, graduation before age 24 or after age 35).¹¹ Missing graduation years are also flagged.¹² Using the seniority variable from the provider, which has no missing values, I impute the flagged values of doctors’ years of experience. I impute their years of experience in the following way: eight years for doctors who have completed specialist training, five years for those who are in specialist training, and two years for medical doctors without specialization. I include a variable flagging imputed values in all regressions where years of experience is used.

F Robustness

1 Robustness of risk variables and complementarity functional form

1.1 Nonparametric investigation of the functional form of the skill-risk interaction

To investigate the functional form of the interaction effect between patient risk (in terms of lagged outcomes) and standardized doctor value added for the three outcomes, I display binscatter plots in Figure A5. The y-axis represents the negative outcome after the consultation, for instance, Emergency Room, and the x-axis represents the standardized skill of the doctor. The results are plotted separately for four groups, containing equal intervals of the respective patient risk measure, to assess the potential linearity of the interaction effect.

The first point to note is that the negative outcome is always higher for higher-risk groups: for all three outcomes, the lagged outcome risk proxy performs well. Second, for Counter-Guideline and Emergency Room outcomes, the lines are always downward sloping, indicating that patients who are conditionally randomly assigned to higher-skilled doctors experience fewer negative outcomes after the consultation.

Finally, the slope of the reduction in negative outcomes for higher-skilled doctors becomes steeper as patient risk increases, for both Counter-Guideline and Emergency Room outcomes (moving from blue, to red, to green, to orange lines), which are the two outcomes where there exists

¹¹The graduation year can be an incorrect measure of a doctor’s medical school graduation, as it captures the most recent degree completed and may instead reflect a later university degree. In some cases, delays in updating the registry could instead mean that the record year reflects the doctor’s high school graduation.

¹²There are several reasons why the doctor graduation year and medical school variables can be missing. First, doctors do not appear in the register if they are not living in Sweden during the entire year of measurement or have died. Second, they may have received their education abroad. For instance, in 2013, more than half of the MD licenses that were newly recognized that year in Sweden were awarded abroad (991 in Sweden and 1255 abroad, of which 1,052 were in an EU-EES country) (National Board of Health and Welfare(Socialstyrelsen) 2014).

complementarity in my main analysis. Moreover, the slope difference seems to increase by reasonably similar amounts across equally sized intervals of patient risk, which suggests that modeling the complementarity through an interaction of two linear terms is a plausible assumption.

1.2 Robustness of risk variable choice

In the main analysis, I use a parsimonious proxy for patient risk based on a lagged variable for each outcome. The benefit of this approach is that it is easy for healthcare providers to implement in practice, as information about lagged outcomes is available in patients' medical records. Moreover, using one-dimensional risk variables leads to a transparent policy rule. However, there are more precise ways of predicting patient risk.

I explore a more precise way of predicting patient risk using the least absolute shrinkage and selection operator (LASSO). I allow the LASSO to select from many aspects of patients' past healthcare history and demographics, and in one version also current symptoms. This approach leads to better prediction of post-consultation negative outcomes than when using a single lagged outcome alone, in terms of the correlation between patient risk and the patient's post-consultation outcomes, or in terms of the root mean square error (RMSE). However, the differences are not large enough to merit using this more complicated risk prediction in the entire paper, given the benefits of parsimonious risk proxying discussed above. Below, I describe the alternative risk prediction approach in more detail.

LASSO risk prediction. Using a logit LASSO, I predict patients' risk for the three outcomes. The LASSO selects variables from a rich set of patient characteristics, using Stata's default options. Table A23 shows both the set of variables LASSO could select from and those it actually selected. Three classes of variables are included: patients' past medical history, demographic and socioeconomic information, and pre-consultation self-reported symptoms. Past medical history and pre-consultation self-reported symptoms are, in theory, accessible to any healthcare provider that wants to implement similar methods. Some demographic information, like location and gender, is available to healthcare providers, while socioeconomic information (e.g., income level and immigration status) is mostly unavailable.

Another important consideration in choosing the main risk proxies (simply lagged variables), as I have done in the main text, is that they reflect slow-moving averages over a relatively long period (three years) before the visit. This means that a patient would have a high probability of matching with the same doctor if the matching procedure were implemented repeatedly over time, improving continuity of care. By contrast, some of the variables I add to the more detailed risk prediction in this section are events only in the past 30 days, meaning there would be a higher likelihood of being matched with different doctors over time. A patient's pre-consultation self-reported symptom is even more specific to the visit, which is why I attempt to avoid using it even in this alternative risk prediction section.

I do not let LASSO choose from pre-consultation self-reported symptoms for predicting Primary

Care or Emergency Room outcomes, since I can improve the prediction in terms of correlation between the risk variable and the future outcome only using the health, demographic and socioeconomic variables, and because adding symptoms does not considerably improve either correlation or RMSE for these outcomes. However, as Table A24 shows, for Counter-Guideline outcomes, using only additional health and demographics variables does not improve the correlation between risk and future outcomes. In this case, adding symptoms considerably improves the correlation between the prediction and the outcome, so I also allow the LASSO to select from patients' self-reported symptoms for Counter-Guideline outcomes.

Table A24 shows that using the LASSO model leads to a better prediction than when using the standard lagged outcomes. For all three outcomes, when risk is estimated using LASSO on past healthcare use and demographic variables, either the correlation between the patient's estimated risk and outcome increases or the RMSE decreases, or both. However, the gains are not always large. The most substantial improvement, over lagged outcomes, occurs for the model predicting Counter-Guideline outcomes, when patients' self-reported symptoms are included. In the following section, I use this more detailed risk prediction in complementarity analysis and average reallocation effects.

1.3 Semiparametric complementarity using LASSO-predicted risk

Using the more precise risk prediction, I explore an alternative functional form of the complementarity between patient risk and doctor skill. Specifically, I divide LASSO-predicted patient risk into deciles, for a less parametric functional form, and interact them with linear skill in regressions with future Counter-Guideline prescriptions, Primary Care visits, or Emergency Room visits as outcomes (Appendix Table A25).¹³

There is a close to (up to statistical precision) monotonic relationship between risk deciles and negative outcomes for all three outcomes – the higher the predicted risk decile, the more adverse outcome a patient has after the consultation. Moreover, for Counter-Guideline and Emergency Room outcomes (the two with a significant interaction effect in Table 6), there is a decreasing trend (up to statistical precision) in the interaction effects between doctor skill and patient risk deciles as risk increases.¹⁴

Using these predicted risk deciles instead of standard patient risk based on lagged outcomes in counterfactual reallocations, the average reallocation effect for Counter-Guideline outcomes is -20% (s.e. 1.9), and it is -2.8% (s.e. 1.1) for Emergency Room outcomes, while there is still no significant effect for Primary Care outcomes (Figure A4). The similarity of the average reallocation effects to the main results in Figure 3 for Primary Care and Emergency Room outcomes is reassuring, as it suggests the results are robust. For these outcomes, this suggests that the choice of the risk variable and the use of a linear specification in the main analysis were not consequential.

The larger average reallocation effect for Counter-Guideline outcomes compared to Figure 3

¹³Also using deciles for skill would make the cells too small with regards to statistical power.

¹⁴For Counter-Guideline, the majority of the interaction terms are significant, but for Emergency Room, only the highest two risk deciles are.

stems from the much better prediction of patient risk when symptoms are included. There is an interesting trade-off between patient-doctor continuity and achieving the best patient-doctor match, which is not the case for the other outcomes where symptoms were not included in the risk prediction. The average reallocation effect for Counter-Guideline outcomes used in the paper is the more conservative estimate.

2 Robustness exercise for both value added and complementarity

The robustness checks above explored the choice of the risk variable and the functional form of the complementarity estimation. As an additional robustness check, I follow an exercise in Adhvaryu et al. (2024) (C.1.1 in their Supplemental Appendix), which probes both the doctor skill estimation (value added) and the complementarity estimation in a nonparametric way. This exercise avoids calculating value added completely and instead provides another way of calculating doctor skill. The idea is to calculate doctor types based on patient outcomes in a holdout sample and then use these to estimate the production function (including complementarity) in the rest of the data. Patient types are calculated as in the main paper, with lagged outcomes using healthcare data from three years prior to the online consultation. The exogeneity assumption in the assignment of patients to doctors means that mean patient outcomes in part of the data can be used to recover measures of doctor skill.

I follow Adhvaryu et al. (2024) and draw the first third of each doctor’s time periods (defined as two-week periods) as a holdout sample to define doctor type.¹⁵ I calculate the mean outcomes for each doctor in each outcome in this holdout dataset. To make the results more readable and comparable with the main paper (where the doctor variable is interpreted as skill, i.e., it measures something positive), I multiply the mean negative outcome per doctor from the holdout sample with -1 . I then divide doctors into quartiles of each skill, where a higher quartile corresponds to a higher skill, as in Adhvaryu et al. (2024). Because the risk for Counter-Guideline outcomes contains many zeros, it cannot be divided into quartiles but only into terciles. Hence, I divide all patient risks (the regular lagged outcomes) into terciles for consistency. Then, for the complementarity estimation, I interact patient risk terciles with doctor skill quartiles in the remaining part of the sample, having removed the holdout sample.

The results are displayed in Appendix Table A26. There are 155,777 observations left after removing the first 33% of doctors’ time periods (not exactly one-third of the total dataset’s observations, as observation numbers vary across different two-week time periods). There are still 435 doctors included, as in the main analysis, and the outcome means for this remaining sample are very similar to those in the full sample. Standard errors are clustered at the doctor level as in Adhvaryu et al. (2024).

For all three outcomes, patients in risk terciles 2 and 3 have higher post-visit adverse outcomes than the omitted first tercile, and the third tercile has the highest adverse outcomes. This is

¹⁵Adhvaryu et al. (2024) used the first three months, corresponding to a third of the total data. However, because the doctors in my setting work with different frequencies and often part time, I draw the first third of each doctor’s time periods.

consistent with the results in the main paper showing that the lagged outcomes work well as a measure of patient risk.

The two highest quartiles in terms of doctor skill (here defined as doctors who have lower mean adverse outcomes in the holdout sample) have significantly lower post-consultation adverse Counter-Guideline and Primary Care outcomes among low-risk patients (first tercile), consistent with these doctor skills also benefiting low-risk patients. For Emergency Room outcomes, there is no significant effect of doctor skill on low-risk patients.

The interaction effects between patient risk terciles and doctor skill quartiles are generally negative (compared to the lowest risk and lowest skill group) for Counter-Guideline and Emergency Room outcomes, and the interaction terms for Emergency Room tend to become more negative as both risk and skill increase, up to statistical precision. For Primary Care outcomes, no interaction effects are found, consistent with the main paper’s findings.

The interaction effects for risk and skill are generally insignificant for Counter-Guideline outcomes, but are significant for Emergency Room outcomes starting from the third risk tercile and third skill quartile. In sum, the results are consistent with those in the paper, though effects are less significant, which could partly be due to the smaller sample size.

3 Robustness on doctor volume

The interaction effects are robust to restricting to doctors with high volume. Table A4 explores the robustness of the estimates shown in Table 6 to exclude doctors with few observed consultations. The analysis sample of 435 doctors is restricted here to the 218 doctors who have weakly above the median number of within-sample consultations; the median is 206.

The main results are also robust to restricting to high-volume doctor time periods. I have added the sample restriction that each observed doctor two-week time-period cell must contain at least 30 observations. Then, I have added the restriction again that each doctor must be observed in at least two time periods. This leaves 172,177 observations to estimate value added and 233 doctors (sample for Table 4). As usual, the first period will not get a value added, so the observations for any following analysis will be 161,135 (sample for Table A3). The results in Table A3 are very similar to those in Table 6.

G Estimation of the distribution of doctor skills

I estimate the distribution of doctor skills following Rose, Schellenberg, and Shem-Tov (2022) and Kline, Saggio, and Sølvssten (2020). First, I residualize the outcomes on the controls of patients with whom a doctor met within time period t , which in my setting is a two-week period.

Let the doctor mean residuals by time period, $\bar{Y}_{jt}$, be defined as follows:

$$\bar{Y}_{jt} = \frac{1}{n_{jt}} \sum_{i|(j(i,t)=j)} (Y_{it} - X'_{it}\hat{\Gamma}) = a_j + \bar{v}_{jt}, \quad (\text{A14})$$

where n_{jt} is the number of patients assigned to doctor j in time t , $\bar{v}_{jt} = \frac{1}{n_{jt}} \sum_{i|(j(i,t)=j)} (u_{it} + X'_{it}(\Gamma - \hat{\Gamma}))$, u_{it} is defined in Equation 2, and $\mathbb{E}[\bar{v}_{jt}] = 0$. I assume that $\bar{v}_{jt}$ is uncorrelated across time periods:

Assumption 2 (uncorrelated doctor-time period estimation error): $\mathbb{E}[\bar{v}_{jt}\bar{v}_{jt'}] = 0 \forall j, t \neq t'$

For different outcomes, A, B , one can estimate the covariance (and variance when $A=B$) of the doctor effects across outcomes using the following estimator:

$$\widehat{\text{Cov}}(a_j^A, a_j^B) = \sum_i \sum_{k \neq i} C_{ik}^{AB} Y_i^A Y_k^B, \quad (\text{A15})$$

$$C_{ik}^{AB} = \begin{cases} \frac{J-1}{J^2} \frac{1}{|T_j^A| |T_j^B| - |T_j^A \cap T_j^B|} & \text{if } j(i) = j(k) \\ \frac{-1}{|T_{j(i)}^A| |T_{j(k)}^B| J^2} & \text{if } j(i) \neq j(k) \end{cases} \quad (\text{A16})$$

where i indexes the doctor-date mean residuals with $Y_i := \bar{Y}_{j(i)t(i)}$. J is the total number of doctors, which is constant across outcomes. Finally, I define T_j^A as the set of dates in which doctor j is observed with a patient defined for outcome A . Standard errors for the estimates in Equation A15 are computed analytically from the implied sampling covariance matrix. The expression below gives the sampling covariance conditional on the latent doctor effects; in implementation, unknown quantities are replaced by their plug-in (leave-out) estimates estimates when computing analytic standard errors, following Kline, Saggio, and Sølvesten (2020).

$$\begin{aligned} & \text{Cov}(\widehat{\text{Cov}}(a_j^A, a_j^B) - \text{Cov}(a_j^A, a_j^B), \widehat{\text{Cov}}(a_j^C, a_j^D) - \text{Cov}(a_j^C, a_j^D)) \\ &= \sum_i \sigma_i^{AC} \left(\sum_{k \neq i} C_{ik}^{AB} a_{j(k)}^B \right) \left(\sum_{k \neq i} C_{ik}^{CD} a_{j(k)}^D \right) \\ &+ \sum_i \sigma_i^{AD} \left(\sum_{k \neq i} C_{ik}^{AB} a_{j(k)}^B \right) \left(\sum_{k \neq i} C_{ik}^{CD} a_{j(k)}^C \right) \\ &+ \sum_i \sigma_i^{BC} \left(\sum_{k \neq i} C_{ik}^{AB} a_{j(k)}^A \right) \left(\sum_{k \neq i} C_{ik}^{CD} a_{j(k)}^D \right) \\ &+ \sum_i \sigma_i^{BD} \left(\sum_{k \neq i} C_{ik}^{AB} a_{j(k)}^A \right) \left(\sum_{k \neq i} C_{ik}^{CD} a_{j(k)}^C \right) \\ &+ \sum_i \sigma_i^{AD} \sum_{k \neq i} C_{ik}^{AB} C_{ik}^{CD} \sigma_k^{BC} + \sum_i \sigma_i^{AC} \sum_{k \neq i} C_{ik}^{AB} C_{ik}^{CD} \sigma_k^{BD}. \end{aligned} \quad (\text{A17})$$

where $\bar{Y}_{j(i)t(i)}^k = a_{j(i)}^k + \bar{v}_{j(i)t(i)}^k$ and σ_i^{AB} is defined as the covariance between $\bar{v}_{j(i)t(i)}^A$ and $\bar{v}_{j(i)t(i)}^B$.

The sample variances are a special case of the covariance formula above. Applying the delta

method, analytic standard errors for the estimates presented in Table 4 are calculated.

H Tests of the value-added method

The standard tests of the value-added method involve evaluating the forecast bias and inspecting the autocorrelations, as in Chetty, Friedman, and Rockoff (2014).

1 Forecast bias

Forecast bias is defined by Chetty, Friedman, and Rockoff (2014) using a hypothetical experiment in which patients are randomly assigned to doctors (in their study, students are not conditionally randomly assigned to teachers). λ is defined as the mean impact of being randomly assigned to a doctor who is 1 unit higher in value added, based on observational data from previous years. Then, forecast bias in a value-added model is defined as $B = 1 - \lambda$. Value-added estimates from an observational setting are forecast unbiased if $B = 0$, which means that doctors who have a one-unit higher estimated value added in fact cause patients' outcomes to improve by one unit on average. Chetty, Friedman, and Rockoff (2014) develop tests for forecast bias. One of them, which is used in the literature and feasible in my setting, is to estimate forecast bias based on the degree of selection on observable characteristics that are excluded from the value-added model. This is feasible in my setting because I have additional patient characteristics that are not included in the baseline model, as I want the matching procedure to be implementable using the more limited information that healthcare providers would have access to.

In particular, I generate predicted outcomes for each patient based on the following characteristics: patient total income measured in 2015, where missing values are set to 0; an indicator for whether the income data are missing, e.g., for children; indicators for first- and second-generation immigrant status; and the number of specialist care visits in the three years prior to the online consultation. All these patient characteristics are significantly related to at least one of the outcomes, and many of them are significantly related to all of the outcomes. Table A23 also shows that similar variables are often selected by LASSO in a more detailed risk prediction than the one used in the main paper.

I choose the socioeconomic variables because I do not use socioeconomic variables in the main version of the analysis, since they are not observable to a healthcare provider who would be implementing matching. At the same time, there is a socioeconomic gradient in health and healthcare use also in Sweden, which is why one would expect these variables to be relevant (Chen, Persson, and Polyakova 2022). Furthermore, I add one additional health-related variable because, as including something directly related to health provides a stronger test for forecast bias. The variable is the number of past specialist care visits, chosen as I do not use it in the rest of the analysis, as it is not a direct lagged outcome of the outcome variables.

The results, presented in Table 3, are discussed in the main paper. The estimated forecast bias can be seen from Column 2: it is 0.056%. The upper bound of the 95% confidence interval is 0.53%

– calculated as $(0.00056 + 1.96 \cdot 0.0024) \cdot 100$ – which is in the same order of magnitude as the upper bound of the first forecast bias estimate in Chetty, Friedman, and Rockoff (2014, Table 3) of 0.25%.

2 Autocorrelations

Figure 2 plots the autocorrelations for each of the three mean residualized outcomes. The autocorrelations are between 0.2 and 0.05 in the first three periods, depending on the skill, and then gradually become lower, as expected. After around 15 periods, the lines flatten out, just as in Chetty, Friedman, and Rockoff (2014) (in their case after 7 periods), and this is where I set the maximum drift limit. The autocorrelations are lower for the outcomes that I observe in the primary care setting compared to the outcomes in Chetty, Friedman, and Rockoff (2014) (test scores), but not by several orders of magnitude. In Chetty, Friedman, and Rockoff (2014), the autocorrelations for English range between 0.1 and 0.3 in the first three periods, and for math they range between 0.3 and 0.5 in the first three periods, after which they gradually become lower. The autocorrelations for English in middle school in Chetty, Friedman, and Rockoff (2014) go down to approximately 0.05 at the drift limit, which is only slightly higher than the lowest autocorrelations (closest to the drift limit) in this paper, which fluctuate around 0.03.

References for Appendix

- Adhvaryu, Achyuta, Vittorio Bassi, Anant Nyshadham, and Jorge Tamayo. 2024. “No Line Left Behind: Assortative Matching Inside the Firm”. In: *The Review of Economics and Statistics*, 1–45. ISSN: 0034-6535. DOI: 10.1162/rest_a_01516. eprint: https://direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01516/2477299/rest_a_01516.pdf.
- Anell, Anders, Margareta Dackehag, Jens Dietrichson, Lina Maria Ellegård, and Gustav Kjellsson. 2024. “Better off by risk adjustment? Socioeconomic disparities in care utilization in Sweden following a payment reform”. In: *Journal of Policy Analysis and Management*, 1–33. DOI: <https://doi.org/10.1002/pam.22610>.
- Anell, Anders, Jens Dietrichson, Lina Maria Ellegård, and Gustav Kjellsson. 2023. “Well-Informed Choices? Effects of Information Interventions in Primary Care on Care Quality”. Unpublished manuscript. DOI: /10.2139/ssrn.4441923.
- Bau, Natalie. 2022. “Estimating An Equilibrium Model of Horizontal Competition in Education”. In: *Journal of Political Economy* 130 (7): 1717–1764. DOI: <https://doi.org/10.1086/719760>.
- Bendz, Anna. 2011. “Vårdvalet i Västsverige”. In: *Västsvensk vardag*. Ed. by Annika Bergström. Göteborgs universitet: SOM-institutet, pp. 71–79.
- Chen, Yiqun, Petra Persson, and Maria Polyakova. 2022. “The Roots of Health Inequality and the Value of Intrafamily Expertise”. In: *American Economic Journal: Applied Economics* 14 (3): 185–223. DOI: 10.1257/app.20200405.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates”. In: *American Economic Review* 104 (9): 2593–2632. DOI: 10.1257/aer.104.9.2593.
- Ekonomifakta. 2023. *Regional statistik*. Accessed February 26, 2025. URL: <https://www.ekonomifakta.se/regional-statistik/ditt-lan-i-siffror/skane/>.

- Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey. 1998. "Comorbidity measures for use with administrative data". In: *Medical Care* 36 (1): 8–27. DOI: <https://doi.org/10.1097/00005650-199801000-00004>.
- Ellegård, Lina Maria, Gustav Kjellsson, and Linn Mattisson. 2022. "An App Call a Day Keeps the Patient Away? Substitution of Online and In-Person Doctor Consultations Among Young Adults". Mimeo.
- Frandsen, B., L. Lefgren, and E. Leslie. 2023. "Judging Judge Fixed Effects". In: *American Economic Review* 113 (1): 253–277. DOI: 10.1257/aer.20201860.
- Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-Out Estimation of Variance Components". In: *Econometrica* 88 (5): 1859–1898. DOI: <https://doi.org/10.3982/ECTA16410>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16410>.
- Miller, Amalia R and Catherine E Tucker. 2011. "Can Health Care Information Technology Save Babies?". In: *Journal of Political Economy* 119 (2): 289–324. DOI: <https://doi.org/10.1086/660083>.
- National Board of Health and Welfare(Socialstyrelsen). 2014. *Statistik om hälso- och sjukvårdspersonal: Officiell statistik om antal legitimerade (2013) och arbetsmarknadsstatus (2012)*. URL: <https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/statistik/2014-9-8.pdf>.
- Nationell Patientenkät (Socialstyrelsen). 2016. *The National Patient Survey*. Accessed January 1 2019. URL: <https://resultat.patientenkät.se>.
- Nationell Patientenkät (Socialstyrelsen). 2018. *The National Patient Survey*. Accessed January 1 2019. URL: <https://resultat.patientenkät.se>.
- NYU. 2023. *Commonwealth Fund-Supported Analysis of New York City Electronic ED Records by the NYU Center for Health and Public Service Research and the United Hospital Fund of New York*. Accessed August 7, 2023. URL: www.wagner.nyu.edu/faculty/billings/nyued-background.
- Olsson, Anna-Clara, Annica Isacson, and Lars Kolmodin. 2022. *Läkartillgång i primärvården*. Technical Report. Accessed December 13, 2023. Sveriges Kommuner och Regioner. URL: https://extra.skr.se/download/18.1c1463c417f1b32c7ab447ab/1646054902587/L%C3%A4kare%20i%20prim%C3%A4rv%C3%A5rden_SKR.pdf.
- Region Skåne. 2017. *Care Need Index and Lipid Lowering Treatments*. Accessed January 2, 2017. URL: <https://vardgivare.skane.se>.
- Rose, Evan K, Jonathan Schellenberg, and Yotam Shem-Tov. 2022. "The Effects of Teacher Quality on Adult Criminal Justice Contact". Mimeo.
- Stagg, Vicki, Robert Hilsden, and Hude Quan. 2015. *ELIXHAUSER: Stata module to calculate Elixhauser index of comorbidity*. Statistical Software Components S458077, Boston College Department of Economics.
- Strama. 2017. *Rekommendationer för kvalitetsindikatorer vid digitala vårdmöten*. Replaced by 2022 version. URL: <https://strama.se/wp-content/uploads/2025/04/kvalitetsindikatorer-for-digitala-vardmoten-2022.pdf>.
- Strama. 2019. *Rekommendationer för kvalitetsindikatorer vid digitala vårdmöten*. Replaced by 2022 version. URL: <https://strama.se/wp-content/uploads/2025/04/kvalitetsindikatorer-for-digitala-vardmoten-2022.pdf>.
- Tillväxtverket. 2011. *Tillgänglighet till kommersiell och offentlig service 2011*.