Web Appendix

'The Effectiveness of Health Screening', *Health Economics*, 201x, Vol. **x**(x), xxx-xxx.

Franz Hackl[*], Martin Halla[†], Michael Hummer[*], and Gerald J. Pruckner[*]

Last update: May 8, 2014
(First version: September 24, 2013)

Abstract

This document provides detailed estimation output for the robustness tests discussed in Section 6 of the paper.

[*]University of Linz
[†]Corresponding author: University of Linz and IZA; email: martin.halla@jku.at

Introduction

Section 6 of the paper discusses briefly robustness checks which should help to dispel concerns regarding the validity of the identifying assumption of our instrumental variable approach. The plausibility of the instrumental variable (henceforth IV) can be questioned based on two related grounds:

- The positive correlation between the individual propensity to participate in health screening and the average screening rate in a given zip code (i.e., our first stage relationship) might be the result of peer effects and is not driven by the supply side as our identification strategy presumes. If this is true, and iff equivalent peer effects are also present in health-care utilization, sick leave, and hospitalization (our second stage relationships), then the identifying assumption of our IV strategy is not fulfilled.
- Second, sorting of insurants into certain zip-code areas might invalidate our IV strategy if this is correlated with unobserved confounding factors. In particular, one might question whether a general health awareness or tendency towards preventive health care activities at the zip-code area level explains the utilization of screening examinations and whether this also matters in our second stage relationships.

Obviously, given the nature of the IV method, one cannot provide watertight proof for the validity of our identifying assumptions. In the following, however, we will present comprehensive and, from our point of view, convincing evidence for each of the above mentioned concerns. We hope that this evidence will convince critical readers of the reliability of our results. We also want to emphasize the major identification challenge evident in the evaluation literature of large-scale screening programs. We believe that our estimation strategy constitutes an important step towards solving the endogeneity problem and provides quite reliable estimates of the effects of screening.

Peer Effects

Clearly there is no direct test for the existence of peer effects in the screening decision. However, we offer a falsification test providing suggestive evidence against peer effects. This test uses the following logic: The existence of peer effects would imply that the correlation in screening behavior is higher among members of a peer group (*intraclass correlation*) compared to the correlation between members and non-members. Consequently, average screening rates should vary considerably across different peer groups.[1] In contrast, the proposition put forward by our IV strategy — the supply-side determined screening rates — would predict similar average screening rates across different peer groups within a zip-code area, as they are all exposed to the same supply side. Put differently, this would imply a high correlation between screening rates of different peer groups within a zip-code area. Thus, one could formulate the following simple empirical test:

[1] To be precise, it is theoretically possible that all peer groups have the same average screening rate just by chance.

- Evidence for peer effects: There is no (or a low) correlation between average screening rates across peer groups within a zip-code area.
- Evidence for supply-side determination: There is a high correlation between average screening rates across peer groups within a zip-code area.

To implement this test, one has to define potential peer groups (i. e., individuals who are likely to meet and mutually influence each other's beliefs and behavior). While our administrative data do not include information on personal links between insurants, we are able to use personal characteristics known to determine personal links, such as friendships and acquaintanceships.[2] We suggest the following dimensions to implement our test:

- Religion (Catholic vs. non-Catholic)
- Ethnicity (Austrian vs. non-Austrian citizens)
- Educational attainment (academic degree vs. no academic degree)
- Income (first quartile vs. third quartile)
- Industry of employer (production worker vs. non-production worker)
- Wage earners vs. self-employed

We assume here that people with the same religious denominations and ethnicity, with similar educational attainment, within the same income range, or those employed in the same industry are more likely to meet and mutually influence each other's beliefs and behavior regarding screening, compared to people from different groups defined by these criteria. The final definition of peer groups — along the lines of wage earners versus self-employed individuals — is particularly interesting, as these groups are also covered by a different mandatory health insurance fund. Note: insurants across all peer groups residing in the same zip-code area face the same local supply side.

Table 1: Correlation between average screening rates of different groups

Peer groups	Correlation between avg. screening rates
Religion: Catholic vs. non-Catholic	0.730***
Ethnicity: Austrian vs. non-Austrian citizens	0.933***
Education: academic degree vs. no academic degree)	0.702***
Income: first quartile vs. third quartile	0.873***
Industry of employer: Production vs. non-production	0.807***
Wage earners vs. self-employed	0.733***

Notes: *** indicate statistical significance at the 1-percent level

Table 1 shows the correlation between the average screening rates of the respective groups measured at the zip-code level. It turns out that all correlation coefficients are very high and

[2]A common definition of peer group (see, for instance, Wikipedia) is as follows: '*A peer group may be defined as a group of people who, through homophily, share similarities such as age, background, and social status.*'

highly statistically significant. Therefore, our empirical test provides evidence for the importance of supply-side screening recommendations (and no evidence of peer effects). Put differently, this empirical test supports the validity of our IV strategy.

Potential sorting and other confounders

As a further falsification test, we check the sensitivity of our results to the inclusion of a list of additional covariates measured on the zip-code area level, which constitute potential confounding factors; they can be grouped as follows

- Proxies for a general tendency towards preventive care measures
- Characteristics of GPs' and insurants

In the first step we aimed to test whether our results are confounded by insurants' general health-awareness or tendency towards preventive healthcare activities on a local level. In principle, a number of variables indicate a general tendency towards preventive care measures. However, many are not available in our administrative database (e. g., membership in sports clubs or fitness studios would be a proxy for inclination towards preventive health care measures; individual smoking behavior would test insurant health awareness). Therefore, we focus on the incidence of other preventative healthcare services (other than general screening) and use the average participation rates at the zip-code area level as proxy variables for general health consciousness across regions. In particular, we measure the regional density in the utilization of the following preventive care measures (at the zip-code area level in the year of screening):

- mammography,
- pre- and postnatal mother-child healthcare examinations
- dermatological preventive care examinations (especially for birthmarks).

For these examinations, regular doctor visits are highly recommended. If inclusion of participation rates of these preventive measures changes our screening coefficients substantially, we have evidence that our results are confounded by a general attitude towards preventive healthcare measures on a regional level. On the other hand, stable coefficients would support the plausibility of our IV strategy.

Table 2 summarizes the results of our robustness checks for the most important outcome variable: outpatient expenditures. Column (I) lists results based on our baseline specification. Since mammography examinations are targeted exclusively to women, original estimations of the effects of screening participation are replicated for women only, as can be seen in column (II).[3] Columns (III)-(V) show estimation results if we include the different proxy variables as additional covariates (the respective proxy variable is listed in the header of each column). Each column contains the estimated coefficients of the screening participation for the different lags

[3] Although mother-child preventive care refers, at first glance, to women, we believe that these medical check-ups are family decisions — from the perspective of the four male authors, we appreciate the opportunity to participate in issues of child rearing. Note, however, that results are similar if we observe women only in the estimations.

together with the standard errors and 95% confidence intervals. The estimated coefficients for the included proxies are also listed.[4]

While some of the point estimates change slightly in the extended specifications the different estimations provide consistent results. In each case, we have a large overlap between the 95% confidence intervals from the estimated coefficients of the original model and the extended model. The overlap (measured in percentage) is listed in the table (printed in bold numbers).[5] The high percentage values indicate that our IV estimation is not confounded by a general tendency towards preventive healthcare activities (or specific health-awareness) on a local level. Hence, our testing strategy failed to invalidate the plausibility of our IV strategy.

We also ran a specification where we controlled for the following insurant and GP characteristics (measured at the zip-code area level):

- Insurants
  - Share of female
  - Share of foreign insurants
  - Share of 65 years and older
  - Labor market status distribution
    (Share of retirees and unemployed)
- GPs
  - Share of females
  - Average age
  - University
    (Share of graduates from the universities in Vienna, Innsbruck, Graz, and elsewhere)

The estimation results where we concurrently controlled for these potential health-related confounders are summarized in column (VI). Again, we did not observe considerable changes in the estimated screening coefficients (compared to the baseline specification). With the exception of lag 1, the overlap in confidence intervals of the screening coefficients was beyond 97 percent. We thus conclude that our estimation strategy also is not invalidated by characteristics of the GPs and the beneficiaries.[6]

Finally, following a referee's advice, we also ran a specification (summarised in column (VII)), where we control for the average number of non-screening GP visits (measured at the zip-code level). Again, we observe some changes in the estimated screening coefficients. However, we observe a large overlap between the 95% confidence intervals from the baseline specification and the new extended model.

[4]The proxy variables exhibit, as expected, a positive effect in our second stage estimation. The coefficients provide the estimated effect on outpatient expenditures if the respective participation would jump from zero participation to full participation in mammography, mother-child preventative care, and dermatological examinations. In line with our expectation, we also observe positive effects in the first stage estimations (which are not listed in the table).

[5]The overlap would even be higher if the 99% confidence interval would have been chosen.

[6]Most of the coefficients for the controls (not listed in the table) were statistically significant in both the first and second stage; however, their economic significance is very low.

Table: Robustness analysis of our IV strategy: Effects of potential confounders on a zip-code area level

Lag		Baseline specification (BS)		BS plus additional covariate(s)				
		(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
				Mammography rate[a]	Mother-child preventive care examination rate[b]	Dermatological examination rate[c]	GPs' & insurants' characteristics[d]	Numbers of non-screening GP visits[e]
	Sample:	*all insurants*	*females only*	*females only*	*all insurants*	*all insurants*	*all insurants*	*all insurants*
0	Coef IV	278.68	299.02	290.63	216.21	470.51	284.40	185.14
	SE IV	63.71	83.05	83.36	64.33	341.52	65.37	64.88
	CI 95%	[153.8; 403.6]	[136.2; 461.8]	[127.2; 454.0]	[90.1; 342.3]	[-198.9; 1,139.9]	[156.3; 412,5]	[57.9; 312.3]
	Overlap[f]			**97.6%**	**75.5%**	**100.0%**	**99.0%**	**63.5%**
	Control: Coef (se)			230.09 (127.98)	105.16 (11.16)	-275.80 (444.76)	not listed	6.99 (0.86)
1	Coef IV	195.39	206.40	206.12	176,25	130.66	237.77	149.28
	SE IV	62.54	82.88	82.67	62,67	63.08	64.12	62.64
	CI 95%	[72.8; 318.0]	[44.0; 368.8]	[44.1; 368,2]	[53.4; 299.1]	[7.0; 254.3]	[112.1; 363.4]	[26.5; 272.1]
	Overlap[f]			**99.7%**	**92.3%**	**74.0%**	**84.0%**	**81.3%**
	Control: Coef (se)			364.51 (134.09)	68.19 (14.03)	188.98 (83.89)	not listed	5.5 (0.91)
2	Coef IV	38.47	38.09	44,35	25.28	8.55	40.32	19.07
	SE IV	64.11	83.50	83,16	64.17	63.31	66.78	63.83
	CI 95%	[-87.2; 164.1]	[-125.6; 201.8]	[-118.6; 207.3]	[-100.5; 151.1]	[-115.5; 132.6]	[-90.6; 171.2]	[-106.0; 144.2]
	Overlap[f]			**97.9%**	**94.8%**	**87.5%**	**100.0%**	**92.1%**
	Control: Coef (se)			402.19 (141.71)	66.70 (15.68)	247.30 (90.98)	not listed	4.47 (0.97)
3	Coef IV	-144.56	-198.85	-176,98	-140,62	-139,39	-131.86	-147.63
	SE IV	67.99	86.5	86,39	67,93	67,93	69.91	67.87
	CI 95%	[-277.8; -11,3]	[-368.4; -29.3]	[-346.3; -7.7]	[-273.8; -7.5]	[-272.5; -6.2]	[-268.9; 5.2]	[-280.6; -14.6]
	Overlap[f]			**93.5%**	**98.5%**	**98.0%**	**96.6%**	**98.8%**
	Control: Coef (se)			309.24 (154.61)	45.45 (17.44)	129.97 (97.53)	not listed	3.39 (1.09)

To be continued on next page . . .

. . . *continued from previous page.*

4	Coef IV	-292.84	-325.59	-304.03	-299.15	-281.78	-308.82	-292.31
	SE IV	76.99	95.84	95.87	77.18	77.2	80.49	77.03
	CI 95%	[-443.7; -141.9]	[-513.4; -137.7]	[-491.9; -116.1]	[-450.4; -147.9]	[-433.1; -130.5]	[-466.6; 151.1]	[-443.3; -141.3]
	Overlap[f]			**94.3%**	**98.0%**	**96.0%**	**97.0%**	**100.0%**
	Control: Coef (se)			264.08 (169.07)	53.21 (19.99)	63.18 (108.75)	not listed	3.53 (1.25)
5	Coef IV	-289.44	-360.85	-342.17	-279.20	-274.22	-282.64	-261.59
	SE IV	101.21	119.86	120.92	101.58	101.81	104.17	101.62
	CI 95%	[-487.8; -91.1]	[-595.8; -125.9]	[-579.2; -105.2]	[-478.3; -80.1]	[-473.8; -74.7]	[-486.8; -78.5]	[-460.8; -62.4]
	Overlap[f]			**96.5%**	**97.6%**	**96.0%**	**99.7%**	**100.0%**
	Control: Coef (se)			233.38 (206.48)	23.15 (22.94)	96.35 (125.06)	not listed	3.57 (1.49)
6	Coef IV	-162.52	-182.26	-154.65	-156.73	-157.22	-147.44	-137.02
	SE IV	115.23	129.78	129.90	115.05	115.64	119.29	116.81
	CI 95%	[-388.4; 63.3]	[-436.6; 72.1]	[-409.3; 100.0]	[-382.2; 68.8]	[-383.9; 69.4]	[-381.2; 86.4]	[-366.0; 91.9]
	Overlap[f]			**94.6%**	**98.6%**	**99.0%**	**98.4%**	**100.0%**
	Control: Coef (se)			451.88 (246.98)	22.40 (29.2)	-31.28 (166.21)	not listed	2.11 (1.69)
7	Coef IV	-207.96	-145.92	-161.30	-214.01	-210.71	-208.88	-216.46
	SE IV	136.52	173.65	173.04	137.64	136.44	143.28	136.34
	CI 95%	[-475.5; 59.6]	[-486.3; 194.4]	[-500.5; 177.9]	[-483.8; 55.8]	[-478,1; 56.7]	[-489.7; 71.9]	[-483.7; 50.8]
	Overlap[f]			**97.6%**	**99.3%**	**99.5%**	**100.0%**	**98.3%**
	Control: Coef (se)			-224.47 (338.8)	-6.96 (41.54)	-188.32 (224.41)	not listed	-0.50 (1.84)
8	Coef IV	-297.53	-81.91	-170.66	-314.63	-315.41	-234.13	-320.06
	SE IV	219.48	285.42	290.89	220.29	221.97	240.00	220.04
	CI 95%	[-727.7; 132.7]	[-641.3; 477.5]	[-740.8; 399.5]	[-746.4; 117.1]	[-750,5; 119.7]	[-704.5; 236.3]	[-751.3; 111.2]
	Overlap[f]			**93.0%**	**98.2%**	**98.5%**	**97.3%**	**97.5%**
	Control: Coef (se)			-791.78 (577.04)	393.45 (75.23)	-547.13 (400.32)	not listed	-0.89 (2.22)

Notes: Dependent variable: outpatient expenditures. The F-values in all first stages of the table are well above critical levels. [a] This specification also controls for the annual mammography participation rate at the zip-code area level. [b] This specification also controls for the annual participation rate in pre- and postnatal mother and child healthcare examinations measured at the zip-code area level. [c] This specification also controls for the annual participation rate in dermatological preventive care examinations (especially birthmarks) measured at the zip-code area level. [d] This specification also controls for a number of insurant characteristics (share of females, share of foreign insurants, share of 65 years and older, share of retirees, and share of unemployed) and GP characteristics (share of females, average age, university) at the zip-code area level. [e] This specification also controls for the number of non-screening GP visits at the zip-code area level. [f] The overlap represents the percentage of the estimated confidence interval based on the baseline specification that overlaps with the estimated confidence interval of the respective extended specification.