

孕期空气污染损害了新生儿早期健康

——来自武汉出生队列追踪调查的经验证据

石智雷 周小强

目 录

附录 I 空气污染在空间尺度上的差异	1
附录 II 投影寻踪模型测算新生儿早期健康指数的步骤	2
附录 III 孕期空气污染对生理性特征指标和功能性特征指标的影响.....	3
附录 IV 稳健性检验	4

附录 I 空气污染在空间尺度上的差异

在本文样本中,最早怀孕女性的怀孕时间是2019年3月,最晚分娩女性的分娩时间是2020年12月。这意味着,在本文中,“孕期”的时间跨度是2019年3月—2020年12月。同时,考虑到本文是将最近监测站点匹配给居民家庭,且有的区有1个监测站点、而有的区有多个监测站点,因此本文以环境监测站点为基本单元,比较了2019年3月1日—2020年月12月31日武汉市各环境监测站点的PM_{2.5}浓度差异。

具体而言,本文测算了各环境监测站点在上述时间范围内的PM_{2.5}浓度均值、PM_{2.5}在轻度污染及以上(PM_{2.5}≥75μg/m³)、中度污染及以上(PM_{2.5}≥115μg/m³)以及重度污染及以上(PM_{2.5}≥150μg/m³)的天数。

表11汇报的是针对22个环境监测站点空气污染的描述统计。从PM_{2.5}浓度均值来看,最小值为41.95μg/m³,最大值为64.91μg/m³,二者相差22.96μg/m³。从PM_{2.5}轻度污染及以上天数,各环境监测站点中,最小值为47天,最大值为197天,二者相差150天。从PM_{2.5}中度污染及以上天数,最小值为4天,最大值为43天,相差39天;从PM_{2.5}重度污染及以上天数看,最大值和最小值相差11天。以上证据表明,各监测站点的PM_{2.5}浓度具有明显差异。

表11 针对各环境监测站点空气污染的描述统计

	均值	标准差	最小值	最大值	样本量
PM _{2.5} 浓度均值(μg/m ³)	50.87	4.49	41.95	64.91	22
PM _{2.5} 轻度污染及以上天数(天)	92.55	30.59	47.00	197.00	22
PM _{2.5} 中度污染及以上天数(天)	20.36	10.95	4.00	43.00	22
PM _{2.5} 重度污染及以上天数(天)	4.95	2.73	0.00	11.00	22

附录 II 投影寻踪模型测算新生儿早期健康指数的步骤

借鉴相关研究的做法（张吉鹏和卢冲，2019；石智雷和滕聪波，2023），本文使用投影寻踪模型测算新生儿早期健康指数的步骤如下：

第一，对新生儿早期健康指标体系进行归一化处理。该步骤是为消除各指标的度量单位和取值范围的影响。设定第 i 个新生儿的第 j 个三级指标为 x_{ij} ($i=1, 2, \dots, n$; $j=1, 2, \dots, m$)。考虑到指标体系中，同时存在负向指标和正向指标。正向指标的值越大，新生儿早期健康水平越高；负向指标值越大，新生儿早期健康水平越低。因此，归一化处理时需要分类考虑。正向指标（如 1 分钟内阿氏评分）按公式（II1）处理，负向指标（如早产程度）按公式（II2）处理。公式（II1）和公式（II2）如下：

$$x_{ij}^* = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j}, \quad (\text{II 1})$$

$$x_{ij}^* = \frac{\max x_j - x_{ij}}{\max x_j - \min x_j}. \quad (\text{II 2})$$

在式（II1）和式（II2）中， $\max x_j$ 和 $\min x_j$ 分别表示 j 指标的最大值和最小值。

第二，进行线性投影。将 m 维数据 x_{ij} 投影到 $a = (a_1, a_2, \dots, a_m)$ 上，进而得到一维投影特征值，用 Z_i 表示：

$$Z_i = \sum_{j=1}^m a_j x_{ij}^*, \quad i=1, 2, 3, \dots, n. \quad (\text{II 3})$$

第三，构建投影指标函数，该函数用 $Q(a)$ 表示，表达式如下：

$$Q(a) = S(a)D(a), \quad (\text{II 4})$$

其中， $S(a)$ 为类间距离， $D(a)$ 为类内密集度。 $S(a)$ 可以表示为投影值 Z_i 的标准差，设投影值 Z_i 的均值为 E ，则：

$$S(a) = \left[\frac{\sum_{i=1}^n (Z_i - E)^2}{n-1} \right]^{\frac{1}{2}}. \quad (\text{II 5})$$

类内密集度 $D(a)$ 的表达式如下：

$$D(a) = \sum_{i=1}^n \sum_{j=1}^n (R - r_{ij}) \times u(R - r_{ij}). \quad (\text{II 6})$$

其中， $r_{ij} = |Z_i - Z_j|$ ，表示投影值 i 和投影值 j 之间的距离。 R 和 $u(R - r_{ij})$ 分别为局部散点密度的宽度指标和单位阶跃函数。

第四，求解最优投影方向向量。投影寻踪模型的基本思路是根据最优投影方向向量 a 来构建综合指数。当 $Q(a)$ 取最大值时，向量 a 便是最优投影方向的向量。因此，该向量可以通过求解式（II7）得到，本文采用遗传算法进行求解。

$$\begin{cases} \text{Max} Q(a) = S(a)D(a) \\ \|a\| = 1 \end{cases}. \quad (\text{II 7})$$

第五，构造新生儿早期健康综合指数。把最佳投影方向 a 和归一化处理后的指标代入式（II3），即求得投影值 Z_i 。由于该值反映了新生儿的综合特征，因而本文将定义为新生儿早期健康指数， Z_i 值越大，表明新生儿早期健康越好。

附录 III 孕期空气污染对生理性特征指标和功能性特征指标的影响

表 III.1 孕期空气污染对生理性特征指标和功能性特征指标的影响 (IV-Probit)

变量	生理性特征指标		功能性特征指标
	先天性缺陷 (1)	特殊照料 (2)	睡眠情况 (3)
$PM_{2.5}^{all}$	-0.0020 (0.0024)	0.0136** (0.0064)	-0.0107 (0.0068)
控制变量	是	是	是
出生月份固定效应	是	是	是
地区固定效应	是	是	是
样本量	2721	2721	2721
弱工具变量检验	54.6126	54.6126	54.6126

注：本表报告的是 IV-Probit 模型的平均边际效应。*、**和***分别表示 10%、5%和 1%的显著性水平。括号内为聚类到社区层面的稳健标准误。弱工具变量检验报告的是 Kleibergen-Paap rk Wald F 统计量。

附录 IV 稳健性检验

1. 剔除双胞胎样本

对于双胞胎样本而言,他们在出生时的一些生理健康指标可能会存在异常现象。以出生体重为例,在本文中,单胎新生儿中低体重儿的比例为 3.14%,而双胞胎新生儿中低体重儿的比例达到 16.36%,高出单胎 13.22 个百分点。显然,双胞胎新生儿如此高的低体重儿比例,并非是其早期健康差,而是双胞胎本身所导致的。由于在构建新生儿早期健康指数时考虑了体重因素,因而将双胞胎纳入分析样本,可能会低估早期健康水平。有鉴于此,本部分对双胞胎样本做了剔除处理。从表 IV1 的估计结果看,孕期空气仍会显著降低新生儿早期健康水平。同时,孕期空气污染对新生儿早期健康非线性影响的趋势依然明显。

表 IV1 稳健性检验: 删除双胞胎样本 (2SLS 估计)

变量	Health (1)	Health (2)	Health (3)	Health (4)
$PM_{2.5}^{all}$	-0.0173*** (0.0051)			
$PM_{2.5}^l$		-0.0080*** (0.0025)		
$PM_{2.5}^m$			-0.0268*** (0.0090)	
$PM_{2.5}^s$				-0.0964*** (0.0340)
控制变量	是	是	是	是
出生月份固定效应	是	是	是	是
地区固定效应	是	是	是	是
样本量	2583	2583	2583	2583
弱工具变量检验	52.3696	47.1290	24.6068	21.3592

注: *、**和***分别表示 10%、5%和 1%的显著性水平。括号内为聚类到社区层面的稳健标准误。弱工具变量检验报告的是 Kleibergen-Paap rk Wald F 统计量。变量 $PM_{2.5}^{all}$ 、 $PM_{2.5}^l$ 、 $PM_{2.5}^m$ 和 $PM_{2.5}^s$ 测度的均是整个孕期的污染,因而对应的工具变量均为整个孕期的平均风速 (FS)。下表同此。

2. 控制跨市迁移因素

在基准回归部分,本文验证了怀孕期间的 $PM_{2.5}$ 浓度会对新生儿早期健康产生显著的负向影响。其中, $PM_{2.5}$ 数据来源于武汉市的各环境监测站点。由于本文的研究对象是那些 2020 年出生在武汉的新生儿,理论上这些新生儿的母亲在怀孕期间也应该在武汉。但如果受访家庭是流动人口家庭,那么该家庭中的女性在怀孕期间则可能会回到户籍地养胎。在此背景下,便不能使用武汉市的空气污染情况来衡量受访女性在怀孕期间遭受的空气污染。有鉴于此,本文对分析样本作了进一步限定,即删除夫妻双方都不具有武汉户口的样本。对这类家庭的女性而言,她们更可能在怀孕期间到外地养胎。从表 IV2 呈现的结果看,控制跨市迁移因素后,结论与基准回归一致。

表 IV2 稳健性检验: 控制跨市迁移因素 (2SLS 估计)

变量	Health (1)	Health (2)	Health (3)	Health (4)
$PM_{2.5}^{all}$	-0.0214*** (0.0061)			
$PM_{2.5}^l$		-0.0102*** (0.0031)		

$PM_{2.5}^m$			-0.0339*** (0.0114)	
$PM_{2.5}^s$				-0.1206*** (0.0412)
控制变量	是	是	是	是
出生月份固定效应	是	是	是	是
地区固定效应	是	是	是	是
样本量	2433	2433	2433	2433
弱工具变量检验	50.0443	43.7214	22.6593	20.6813

3.控制市内通勤因素

除跨市迁移可能导致最近监测站点不能很好衡量女性孕期空气污染情况外,市内跨监测站点的通勤也可能产生相同问题。理论上,可能的跨监测站点通勤主要来自上下班。当女性工作地和居住地对应的监测站点不同时,当前使用的距居民家庭最近的监测站点便不能很好地衡量女性孕期的空气污染情况。本部分将回归样本限定在全职妈妈的群体中,该做法能够较好地排除上下班通勤的影响。结果表明,在控制市内通勤因素的影响后,孕期空气污染仍会对新生儿早期健康产生显著的负向影响。

表 IV3 稳健性检验:控制市内通勤因素(2SLS估计)

变量	Health (1)	Health (2)	Health (3)	Health (4)
$PM_{2.5}^{all}$	-0.0190** (0.0077)			
$PM_{2.5}^l$		-0.0093** (0.0040)		
$PM_{2.5}^m$			-0.0271** (0.0110)	
$PM_{2.5}^s$				-0.1507** (0.0741)
控制变量	是	是	是	是
出生月份固定效应	是	是	是	是
地区固定效应	是	是	是	是
样本量	1083	1083	1083	1083
弱工具变量检验	47.505	36.512	33.052	9.675

4.控制疫情因素

本部分进一步考虑每个受访家庭所在行政区的疫情严重程度,以期尽可能控制疫情因素的影响。具体而言,本文参考石智雷和邵玺(2023)的做法,基于2020年1—4月湖北省和武汉市各级卫健部门公开发布的疫情信息,收集了武汉市各区的新冠疫情累计确诊人数。该变量能够反映新冠肺炎疫情的感染风险,累计确诊人数越多,感染风险越高,受访家庭受疫情影响的可能性越大。下表呈现的结果表明,在考虑各区累计确诊人数后,孕期空气污染仍会对新生儿早期健康产生显著的负向影响。同时,累计确诊人数变量的系数基本为负,但在统计意义上不显著。

表 IV4 稳健性检验:控制考虑疫情因素(2SLS估计)

变量	Health (1)	Health (2)	Health (3)	Health (4)
$PM_{2.5}^{all}$	-0.0221*** (0.0065)			
$PM_{2.5}^l$		-0.0108***		

		(0.0035)		
$PM_{2.5}^m$			-0.0310***	
			(0.0098)	
$PM_{2.5}^s$				-0.1222***
				(0.0451)
累计确诊人数	-0.0000	-0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
控制变量	是	是	是	是
出生月份固定效应	是	是	是	是
地区固定效应	是	是	是	是
样本量	2689	2689	2689	2689
弱工具变量检验	38.618	28.441	27.660	16.351

为缓解疫情因素可能产生的影响，本文还针对武汉封城前出生的样本进行了回归，封城前出生的新生儿在其母亲整个孕期均未受到疫情的影响，这能够在较大程度上排除疫情干扰。表 IV5 结果显示，孕期空气污染仍会对新生儿早期健康产生显著的负向影响。

表 IV5 稳健性检验：仅针对武汉封城前出生样本（2SLS 估计）

变量	Health (1)	Health (2)	Health (3)	Health (4)
$PM_{2.5}^{all}$	-0.0192*			
	(0.0112)			
$PM_{2.5}^l$		-0.0068*		
		(0.0040)		
$PM_{2.5}^m$			-0.0234*	
			(0.0135)	
$PM_{2.5}^s$				-0.0867*
				(0.0512)
控制变量	是	是	是	是
出生月份固定效应	否	否	否	否
地区固定效应	是	是	是	是
样本量	214	214	214	214
弱工具变量检验	55.6523	64.4052	58.7983	27.7411

注：武汉是 2020 年 1 月 23 日开始封城，且本文样本中的新生儿都是在 2020 年出生的。因而，对于封城前出生的新生儿，他们的出生月份都是 2020 年 1 月。因此，在仅针对封城前出生样本进行回归分析时，无需再控制出生月份固定效应。同理，控制变量不包括“封城期间怀孕”变量。

5. 替换被解释变量

在前文分析中，本文主要是基于投影寻踪模型来测算新生儿早期健康指数。如果仅是采用这一种方法来测算，可能存在相关结果由偶然因素驱动的风险。为此，本部分进一步参考余典范等（2023）的做法，采用熵值法来测算新生儿早期健康指数（ $Health_SZ$ ），并将之作为被解释变量进行回归。从下表呈现的结果来看，孕期空气污染仍会对新生儿的早期健康产生显著的负向影响。

表 IV6 稳健性检验：替换被解释变量（2SLS 估计）

变量	Health_SZ (1)	Health_SZ (2)	Health_SZ (3)	Health_SZ (4)
$PM_{2.5}^{all}$	-0.0149***			
	(0.0053)			
$PM_{2.5}^l$		-0.0071***		
		(0.0026)		
$PM_{2.5}^m$			-0.0234**	
			(0.0093)	
$PM_{2.5}^s$				-0.0871**
				(0.0363)
控制变量	是	是	是	是

出生月份固定效应	是	是	是	是
地区固定效应	是	是	是	是
样本量	2689	2689	2689	2689
弱工具变量检验	53.0520	44.4159	24.4812	19.2608

6. 仅保留距监测点较近的样本

本文在基准回归部分是根据受访家庭居住小区和武汉市各环境监测站点的经纬度信息，将受访家庭*i*的相关信息与距该家庭最近的环境监测站点*p*的空气污染数据进行匹配。并且，本文还参考 Currie et al. (2009) 的做法，将样本限定为最近监测站点在 10 公里以内的家庭。本部分继续缩短受访家庭距监测站点的距离。理论上，受访家庭距监测站点越近，站点所监测到的数据对受访家庭居住地空气污染程度的代表性越好。具体地，本部分先后将受访家庭和监测点距离限制在 8 公里和 6 公里以内，结果如表 IV7 和表 IV8 所示。可以看到，孕期空气污染仍会对新生儿早期健康产生显著的负向影响。

表 IV7 稳健性检验：仅保留最近距离在 8 公里以内的样本（2SLS 估计）

变量	Health (1)	Health (2)	Health (3)	Health (4)
$PM_{2.5}^{all}$	-0.0210*** (0.0067)			
$PM_{2.5}^l$		-0.0093*** (0.0030)		
$PM_{2.5}^m$			-0.0290*** (0.0101)	
$PM_{2.5}^s$				-0.1052*** (0.0389)
控制变量	是	是	是	是
出生月份固定效应	是	是	是	是
地区固定效应	是	是	是	是
样本量	2495	2495	2495	2495
弱工具变量检验	40.3199	40.5489	24.6700	20.0494

表 IV8 稳健性检验：仅保留最近距离在 6 公里以内的样本（2SLS 估计）

变量	Health (1)	Health (2)	Health (3)	Health (4)
$PM_{2.5}^{all}$	-0.0166* (0.0086)			
$PM_{2.5}^l$		-0.0064** (0.0032)		
$PM_{2.5}^m$			-0.0221* (0.0116)	
$PM_{2.5}^s$				-0.0726* (0.0389)
控制变量	是	是	是	是
出生月份固定效应	是	是	是	是
地区固定效应	是	是	是	是
样本量	2279	2279	2279	2279
弱工具变量检验	24.0157	30.1910	18.6586	16.6119

参考文献

- [1] Currie, J., M. Neidell, and J. F. Schmieder, “Air Pollution and Infant Health: Lessons from New Jersey”, *Journal of Health Economics*, 2009, 28(3), 688-703.
- [2] 石智雷、邵玺, “三孩政策下新冠肺炎疫情对居民生育意愿的影响”, 《人口研究》, 2023 年第 2 期, 第 78—95 页。
- [3] 石智雷、滕聪波, “三孩政策下生育质量研究”, 《人口学刊》, 2023 年第 5 期, 第 1—16 页。
- [4] 余典范、龙睿、王超, “数字经济与边界地区污染治理”, 《经济研究》, 2023 年第 11 期, 第 172—189 页。
- [5] 张吉鹏、卢冲, “户籍制度改革与城市落户门槛的量化分析”, 《经济学》(季刊), 2019 年第 4 期, 第 1509—1530 页。

注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明附录下载出处。