

几种预测方法对尘肺近年患病数预测效果的评价

宁波市卫生防疫站(315010) 肖国兵 潘翠宝 马藻骅

有效地预测尘肺的发病趋势,对制订尘肺防治规划具有重要意义。不同的预测方法应用于同一批数据,可以得到不同的结果。本文拟对灰色数列GM(1,1)、一阶自回归模型和指数曲线模型的预测效果进行评价和探讨。

1 材料和方法

1.1 原始数据:本市尘肺发病在1982年达到最高峰,其后各年发病人数逐渐下降,并稳定在一定的水平上,故选取1982~1992年尘肺历年现患病人数作为原始数据。同时,为更好地评价预测效果,拟将1991和1992年数据不进入模型。

1.2 预测方法:灰色数列GM(1,1)、一阶自回归模型、指数曲线方程等均按有关文献介绍方法建立。

1.3 预测效果评价:采用后验差检验法,即分别计算后验差比值C及小概率误差P,与评价标准(略)相比较,确定预测精度等级。

2 结果

2.1 以1982~1990年数据进行预测

按有关公式分别计算得下列方程:

$$GM(1,1): \hat{y}_1t = 7455.9441e^{0.034679(t-1)} - 7246.9441 - \sum_{j=1}^{t-1} x(t-1)$$

$$\text{自回归模型: } \hat{y}_2t = 103.0352 + 0.6901\hat{y}_2(t-1)$$

$$\text{指数曲线方程: } \hat{y}_3t = 224.2780 \times 1.049511^t$$

按上述三种方法预测结果见表1。

从后验差比值及小概率误差可见,除指数曲线法预测精度属I级(合格)外,GM(1,1)和自回归模型均属I级(好)。不过三种方法外推值与实际值均存在不少差异。

2.2 以1984~1990年数据进行预测

进一步观察1982、1983年数据,可见1982、1983年现患病人数增长幅度明显大于以后各年,为此考虑将1982、1983年数据舍去,重新进行预测,得:

$$GM(1,1): \hat{y}_1t = 9621.2715e^{0.024367(t-1)} - 9340.2715 - \sum_{j=1}^{t-1} x(t-1)$$

$$\text{自回归模型 } \hat{y}_2t = -14.0833 + 1.07574 \hat{y}_2(t-1)$$

表1 三种模型的预测效果比较

| 年份 | 实际人数 | \hat{y}_1t | ϵ_{1t} | \hat{y}_2t | ϵ_{2t} | \hat{y}_3t | ϵ_{3t} |
|---------|------|--------------|-----------------|--------------|-----------------|--------------|-----------------|
| 1982 | 209 | 209 | 0 | 209 | 0 | 235 | -26 |
| 1983 | 249 | 263 | -14 | 247 | 2 | 247 | 2 |
| 1984 | 281 | 286 | -5 | 274 | 7 | 259 | 22 |
| 1985 | 289 | 287 | 2 | 292 | -3 | 272 | 17 |
| 1986 | 296 | 290 | 6 | 305 | -9 | 286 | 10 |
| 1987 | 301 | 297 | 4 | 313 | -12 | 300 | 1 |
| 1988 | 310 | 308 | 2 | 319 | -9 | 315 | -5 |
| 1989 | 325 | 323 | 2 | 323 | 2 | 330 | -5 |
| 1990 | 333 | 333 | 0 | 326 | 7 | 346 | -13 |
| 1991 | 345* | 347 | -2 | 328 | 17 | 363 | -18 |
| 1992 | 352* | 360 | -8 | 329 | 23 | 381 | -29 |
| 后验差比值 C | | 0.1543 | | 0.1822 | | 0.3855 | |
| 小概率误差 P | | 1 | | 1 | | 0.8889 | |

* 1991,1992年数据未进入模型

$$\text{指数曲线方程 } \hat{y}_3t = 271.9925 \times 1.02863^t$$

按新的预测方程进行拟合和预测,结果见表2。

表2 三种模型预测效果比较

| 年份 | 实际病人数 | \hat{y}_1t | ϵ_{1t} | \hat{y}_2t | ϵ_{2t} | \hat{y}_3t | ϵ_{3t} |
|---------|-------|--------------|-----------------|--------------|-----------------|--------------|-----------------|
| 1984 | 281 | 281 | 0 | 281 | 0 | 280 | 1 |
| 1985 | 289 | 287 | 2 | 288 | 1 | 288 | 1 |
| 1986 | 296 | 293 | 3 | 296 | 0 | 296 | 0 |
| 1987 | 301 | 301 | 0 | 304 | -3 | 305 | -4 |
| 1988 | 310 | 313 | -3 | 313 | -3 | 313 | -3 |
| 1989 | 325 | 325 | 0 | 323 | 2 | 322 | 2 |
| 1990 | 333 | 333 | 0 | 333 | 0 | 331 | 2 |
| 1991 | 345* | 342 | 3 | 344 | 1 | 341 | 4 |
| 1992 | 352* | 352 | 0 | 356 | -4 | 351 | 1 |
| 后验差比值 C | | 0.1001 | | 0.1008 | | 0.1368 | |
| 小概率误差 P | | 1 | | 1 | | 1 | |

结果可见,三种方法预测精度均达到I级(好)水平,外推效果也十分理想。

3 讨论

3.1 本文对同一批数据采用三种不同的方法进行预测,结果表明三种预测方法的预测效果有所不同,这与模型的性质本身、对原始数据的要求有关。自回归模型适合变化趋势较稳定的资料,前一次数值对后一次预测影响较大;指数曲线法要求资料呈指数变化,预测时各数值间的直接关系较小。而GM(1,1)克服

了以往模型要求足够样本含量、典型概率分布等缺陷，不过它也要求资料呈单调上升或单调下降，最好呈指数变化。本例1982~1990年尘肺现患病人数中，由于1982、1983年现患病例数变化较其后几年差别为大，不完全符合这些模型要求，这一因素是造成模型预测效果差的重要原因之一，而非完全由于环境因素所致。相比之下，1984~1990年数据更符合模型要求，预测效果大为提高，三种方法结果较为接近。不过从两次预测结果来看，无论是从P、C值还是外推效果，均以GM(1,1)为最优，这与该模型每次预测时均从整个数列整体上考虑（指a、u项），并及时补充新的内容【指 $\sum_{j=1}^{t-1} x(t-1)$ 】项有关，它较前两种模型考虑更全面。

3.2 本文对同一批历史资料，经过适当取舍后，模型预测精度和外推效果均大为提高，这说明进行模型预测时应应对原始资料进行适当取舍，以更符合模型要求。建议：（1）尽量选用最新数据；（2）原始数据一般10个左右，并非越多越好；（3）尽量剔除最早的、又不适合预测模型要求的数据，可采用描图法、进入模型比较预测效果后逐步剔除等方法。通过建立几组数据子集，再选择较合理的数据也不失为一种好方法。此外，选择恰当的统计指标也很重要，如尘肺现患病人数变化相对较平稳，作为预测指标是较理想的。

3.3 本文中介绍的三种方法，均为计算简便、在普通计算器上即可完成，且只要条件符合，均可用于尘肺发病趋势预测。因此我们认为这些方法切实可行，值得进一步推广应用，尤其是GM(1,1)模型。

汞作业工人血清巯基含量变化的调查研究

淄博市职业病防治院 (255067) 王业英 徐强绪 钟云虎 吉祥和

汞作业工人血清巯基(-SH)含量变化及其与接汞时间的相关关系的研究未见系统报道。本文测定了某厂492名汞作业工人的血清-SH含量，并结合尿汞含量与对照组进行了比较和分析。

不同性别血清-SH含量比较 (单位 $\mu\text{mol/L}$)

| | 例数 | | 血清-SH含量 ($\bar{X} \pm S$) | |
|-----|-----|-----|-----------------------------|----------------|
| | 男 | 女 | 男 | 女 |
| 对照组 | 49 | 87 | 179.36 ± 24.09 | 178.65 ± 22.19 |
| 接汞组 | 143 | 349 | 169.28 ± 28.53 | 163.70 ± 23.46 |

1 对象与方法

1.1 对象 某厂接汞工龄超过1年的工人492名(男143名)作为接汞组。选择无毒物接触史的工人136名(男49名)作为对照组，年龄构成与接汞组相似。

1.2 方法 空腹采集静脉血1ml，离心移出血清贮于冰箱中，尽快用比色法测定血清-SH含量。取晨尿用双硫脲比色法测定尿汞含量。

2 结果与分析

2.1 接汞组血清-SH含量均值为 $165.23 \pm 25.36 \mu\text{mol/L}$ ，比对照组 ($179.04 \pm 22.82 \mu\text{mol/L}$) 明显降低 ($P < 0.001$)。接汞组尿汞含量 ($0.060 \pm 0.025 \mu\text{mol/L}$) 高于对照组 ($0.048 \pm 0.0164 \mu\text{mol/L}$)，差异显著 ($P < 0.001$)。通过对接汞组及对照组分别进行不同性别血清-SH含量比较(见下表)，发现对照组血清-SH含量性别间的差异不显著 ($P > 0.5$)，而接汞组血清-SH含量性别之间的差异显著 ($P < 0.05$)。

2.2 接汞组492名工人中具有头痛、头晕、乏力等神经症状及明显齿龈炎的共33例，未发现情绪改变和明显震颤。上述33名工人以下称症状组，余459名汞作业工人下称非症状组。症状组平均接汞工龄为8.7年，血清-SH含量为 $160.66 \pm 21.56 \mu\text{mol/L}$ 与非症状组 ($165.66 \pm 24.73 \mu\text{mol/L}$) 相比差异无显著性 ($P >$

0.05)，症状组尿汞 ($0.063 \pm 0.027 \mu\text{mol/L}$) 与非症状组 ($0.059 \pm 0.030 \mu\text{mol/L}$) 相比无显著性差异 ($P > 0.5$)。

2.3 汞作业工人血清-SH含量与接汞时间的等级相关分析表明，二者之间存在明显的负相关关系，等级相关系数 $r_s = -0.917$ ($P < 0.01$)。汞作业工人各工龄组血清-SH含量与对照组相比，接汞工龄在5年之内的各组段血清-SH含量降低不明显，5年以上各工龄组的血清-SH含量比对照组有明显下降。而汞作业工人的尿汞含量与接汞工龄之间的相关关系不明显 ($r_s = 0.303$, $P > 0.2$)。

3 讨论与小结

该厂车间环境监测资料提示，接汞工人劳动场所空气汞浓度均值为 0.024mg/m^3 。本次调查的结果表明，在此条件下，接汞5年以后血清-SH含量明显下降；而且血清-SH含量与接汞时间之间存在明显的负相关关系。因此血清-SH含量的慢性汞中毒的诊断方面有一定的参考价值。

(本文承周忠华、刘岳文、徐毅、白怀生、毕研君等医师大力协助，鸣谢！)