

加权综合法在铅中毒发病预测中的应用

胡 功 成 张 书 成 李 俊 峰

加权综合法是对两种或两种以上预测结果进行加权合并,使原模型得以修正,从而提高其预测精度的统计方法^[1]。本文拟用此法分析某蓄电池厂 1984年至 1995年铅烟接触水平与铅中毒的发病关系,并与原模型预测结果相比较,以探讨该方法在职业病预测中的应用前景。

1 资料与方法

收集该厂自 1984年以来,化成、格体、总装 3个车间铅烟监测数据 144个,然后用几何均数表示各年铅烟浓度平均值。选择 1984年以后入厂,在上述 3个车间工作 1年以上的岗位工人 234名为观察对象。其中慢性铅中毒患者 29名。铅烟接触水平采用累积接

铅量 (D_i),即接铅时间与年铅烟浓度平均值之和(毫克·年)。

2 结果

2.1 各剂量组铅中毒累积发病率

按铅烟接触水平分组,求出校正观察人数,进而计算各剂量组铅中毒累积发病率,见表 1。

2.2 Logistic回归预测模型

在上述计算结果的基础上采用最小二乘法求得各剂量组上限值和累积发病率的 Logistic回归模型为 $\hat{P}_i = 1 / [1 + \exp(5.1317 - 1.5351D_i)]$,以此方程推算各剂量组曲线预测值 \hat{P}_i 及残差 ϵ_{ii} ,见表 2。

表 1 各剂量组铅中毒累积发病率

剂量组 (毫克·年)	累积观察人数	病例数	到期人数	校正观察人数	未发病率	累积发病率 (P_i)
0.000~	234	1	7	230.5	0.9957	0.0043
0.300~	226	1	30	211	0.9953	0.0090
0.600~	195	4	44	173	0.9769	0.0319
0.900~	147	4	43	125.5	0.9681	0.0778
1.200~	100	1	9	95.5	0.9895	0.0875
1.500~	90	2	26	77	0.9740	0.1112
1.800~	62	2	7	58.5	0.9658	0.1416
2.100~	53	2	12	47	0.9574	0.1782
2.400~	39	2	8	35	0.9429	0.2251
2.700~	29	3	3	27.5	0.8909	0.3096
3.000~	23	4	2	22	0.8182	0.4351
3.300~ 3.600	17	3	14	10	0.7000	0.6046

2.3 直线回归预测模型

以各剂量组上限值和累积发病率为原始数据,求得直线模型为 $\hat{P}_2 = 0.1564D_1 - 0.1203$,以此方程算得各剂量组直线预测值 \hat{P}_2 及残差 ϵ_{2i} ,见表 2。

作者单位: 014030 包头市卫生防疫站

表 2 铅中毒发病率预测值与实际值比较

实际发病率 R	曲线预测值 \hat{P}_i	$\epsilon_i^2 = (P_i - \hat{P}_i)^2$	直线预测值 \hat{P}_i	$\epsilon_i^2 = (P_i - \hat{P}_i)^2$	加权预测值 \hat{P}_i	$\epsilon_i^2 = (P_i - \hat{P}_i)^2$
0.0043	0.0093	0.0000	- 0.0734	0.0060	- 0.0053	0.0000
0.0090	0.0146	0.0000	- 0.0265	0.0013	0.0074	0.0000
0.0319	0.0230	0.0001	0.0205	0.0001	0.0226	0.0001
0.0778	0.0360	0.0017	0.0674	0.0001	0.0415	0.0013
0.0875	0.0558	0.0010	0.1143	0.0007	0.0661	0.0005
0.1112	0.0856	0.0007	0.1612	0.0025	0.0989	0.0002
0.1416	0.1292	0.0002	0.2081	0.0044	0.1431	0.0000
0.1782	0.1904	0.0001	0.2551	0.0059	0.2018	0.0006
0.2251	0.2715	0.0022	0.3020	0.0069	0.2769	0.0027
0.3096	0.3714	0.0038	0.3489	0.0015	0.3674	0.0033
0.4351	0.4835	0.0023	0.3958	0.0015	0.4681	0.0011
0.6046	0.5974	0.0001	0.4427	0.0262	0.5702	0.0012
2.2159		0.0122		0.0571		0.0110

2. 4 加权综合预测模型

此文为同一资料建立的两个单变量预测模型，因此可以认为它们的预测残差 ϵ_{1i} 和 ϵ_{2i} 相互独立，故可加权综合计算

$$W = \sum \epsilon_2^2 / (\sum \epsilon_1^2 + \sum \epsilon_2^2) = 0.8240$$

$$\hat{P} = W \hat{P}_1 + (1 - W) \hat{P}_2 = 0.8240 \hat{P}_1 + 0.1760 \hat{P}_2$$

由此方程算得各剂量组加权综合预测值 \hat{P} 及残差 ϵ_i ，见表 2

2. 5 拟合优度检验

预测模型预测精度可由拟合优度来反映，拟合优度的大小可用相关系数 R^2 和均方残差 MSE 来评价^[2]， R^2 越接近 1，MSE 越小，拟合效果越好，优度愈高。据公式

$$R^2 = 1 - \frac{\sum (P_i - \hat{P})^2}{\sum (P_i - \bar{P})^2}$$

$$MSE = \frac{\sum (P_i - \hat{P})^2}{n} = \frac{\sum \epsilon^2}{n}$$

分别计算上述三种预测模型的 R^2 和 MSE，结果见表 3 显示三种方法 R^2 值大小依次为加权模型 > 曲线模型 > 直线模型，MSE 则相反，表明加权综合模型性能最优。

表 3 几种预测方法拟合优度比较

预测方法	R^2	MSE
曲线预测法	0.9671	0.0010
直线预测法	0.8462	0.0048
加权综合法	0.9704	0.0009

2. 6 预测结果

假若车间铅烟浓度控制在国家现行卫生标准

0.03mg/m³ 之内时，连续接触铅烟 35 年，曲线预测发病率为 2.88%，直线预测为 4.39%，加权预测法为 3.15%，且符合本资料实际情况。

3 讨论

本文在用寿命表法计算出累积发病率后再用 Logistic 回归模型和直线回归模型的基础上使用加权综合预测法对某蓄电池厂三个车间铅烟接触水平和铅中毒发病关系建立预测模型。经拟合优度检验发现，加权综合法 > 曲线预测法 > 直线预测法，说明加权综合法预测精度高于其他两种模型。并用国家现行车间铅烟卫生标准预测其发病率，曲线预测法为 2.88%，直线预测法为 4.39%，加权预测法为 3.15%。

不同的预测方法有不同的适应对象，直线模型适用于变化趋势较稳定的资料，Logistic 模型适用于变化起伏较大符合生物学特征的资料，但各种单一预测法都有其一定的局限性。加权综合法是在其他预测方法的基础上权衡其他模型之利弊，故可扬长避短，减少预测误差，提高其准确性，且其信息利用率高，运算简便。该方法不仅适用于单变量预测模型的合并，也可应用于多变量预测模型的加权综合，是一种具有广泛运用前景的预测方法。

4 参考文献

- 杜琳，何群英. 卫技人员需求量的加权综合预测. 中国卫生统计, 1991, 8 (4): 19
- 陈向春，周惠仙. 尘肺发病的加权综合预测. 中华劳动卫生职业病杂志, 1993, 11 (3): 172