

## 使用线性回归统计方法分析暴露组与健康相关性的系统比较(续完)

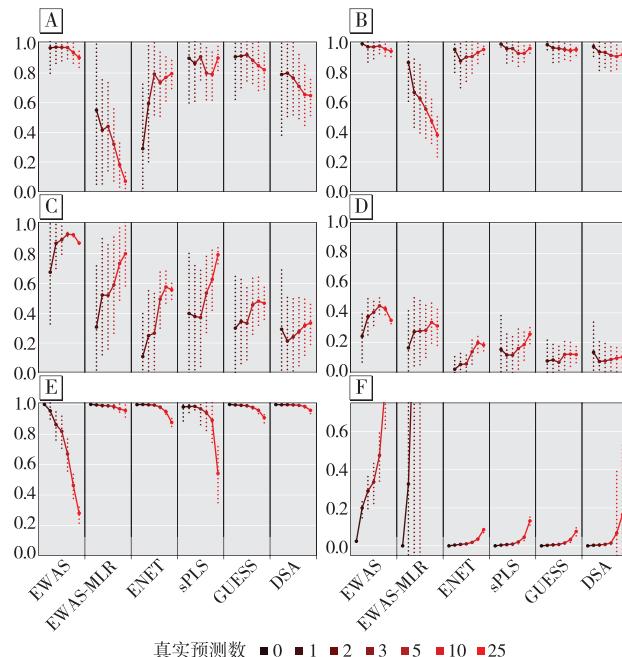
### 3.2 情景组 1 的统计性能评估

情景组 1 的模拟结果见图 1、图 2 和表 1。真实预测变量为完全随机抽取, 单个情景下真实预测变量两两之间的绝对相关性均值(标准误)为 0.12~0.15 (0.12~0.16)。

在所有被分析的真实预测变量数量(即,  $k=0, 1, 2, 3, 5, 10, 25$ )中, EWAS 方法的灵敏度 >90%, 但特异性低至 28%, FDP >67% (见表 1, 由于选择了大量以  $n_B/k$  表示的暴露因素)。各模拟条件下 AltFDP 的范围为 24%~45%。平均绝对偏差较大(范围为 0.02~1.98), 但仅考虑真实预测变量时, 平均绝对偏差在所有统计方法中最小( $\leq 0.10$ , 其他所有方法  $\geq 0.30$ ; 图 S2)。

当 EWAS 后增加一个多元线性回归步骤(EWAS-MLR)时, FDP 在所有情景下均增加(达到 30%~80%), 特异性也提高(所有情景下均 >95%); 但这样改进的代价是灵敏度大幅降低(所有情景下均 <56%)。AltSens 为 38%~87%, AltFDP 为 16%~34%。平均绝对偏差很大(所有情景下平均为 9.00)。

当使用其他校正方法进行多重检验时, 结果是类似的(图 S3)。如果没有应用多重比较的校正, 则改良的 EWAS 得到的 FDP >89%, AltFDP >42%。



[注] 模型的性能以灵敏度(A)、替代灵敏度(AltSens, 见“方法”)(B)、假阳性率(FDP)(C)、替代 FDP(AltFDP, 见“方法”)(D)、特异度(E)和绝对标准偏差(F)表示。在每个被 0~25 个真实预测变量定义的情景下, 运行 100 次得到的统计学指标在图中显示为均值(点), 各统计学指标的变异度由均值两侧的 1 个标准误差(垂直点线)表示。DSA, 删除-替换-加法; ENET, 弹性网络; EWAS, 全环境关联分析; EWAS-MLR, EWAS 多元线性回归分析; GUESS, 图形单元进化随机搜索; sPLS, 稀疏偏最小二乘回归。

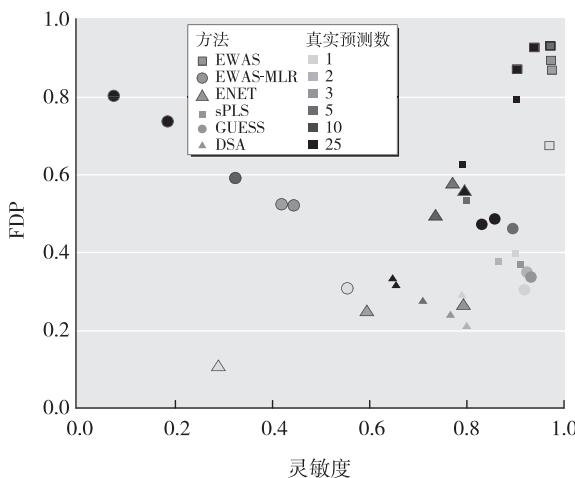
图 1 情景组 1 中各统计学方法的性能

表 1 情景组 1 中统计学方法的统计学效力

方法	灵敏度	AltSens	FDP	AltFDP	特异度	$n_B/k$	平均绝对偏差	TP 的平均绝对偏差
EWAS	0.96[0.90; 0.98]	0.97[0.95; 1.00]	0.86[0.67; 0.93]	0.37[0.24; 0.45]	0.72[0.28; 1.00]	11.27[0.00; 16.66]	0.59[0.02; 1.98]	0.04[0.02; 0.10]
EWAS-MLR	0.33[0.07; 0.55]	0.60[0.38; 0.87]	0.58[0.30; 0.80]	0.27[0.16; 0.34]	0.99[0.96; 1.00]	0.86[0.00; 1.33]	9.00[0.00; 42.72]	0.67[0.45; 0.93]
ENET	0.66[0.29; 0.80]	0.93[0.88; 0.96]	0.37[0.10; 0.57]	0.11[0.02; 0.20]	0.97[0.88; 1.00]	1.15[0.34; 1.87]	0.02[0.00; 0.08]	0.74[0.54; 0.98]
sPLS	0.86[0.79; 0.91]	0.96[0.93; 1.00]	0.52[0.37; 0.79]	0.16[0.12; 0.26]	0.90[0.54; 0.99]	3.59[2.39; 4.78]	0.03[0.00; 0.13]	0.46[0.38; 0.57]
GUESS	0.88[0.82; 0.92]	0.97[0.95; 0.99]	0.39[0.30; 0.48]	0.10[0.07; 0.12]	0.98[0.91; 1.00]	1.45[0.09; 1.79]	0.02[0.00; 0.08]	0.37[0.33; 0.42]
DSA	0.73[0.65; 0.80]	0.94[0.91; 0.98]	0.28[0.21; 0.33]	0.09[0.07; 0.13]	0.99[0.96; 1.00]	0.95[0.26; 1.38]	0.04[0.00; 0.16]	0.51[0.31; 0.89]

[注] 七个情景的结果以均值[最小值, 最大值]显示, 各情景被不同数量的真实自变量定义:  $k=0, 1, 2, 3, 5, 10$  和 25(每个场景模拟 100 次)。

AltFDP, 替代假阳性率(定义见“方法”); AltSens, 替代灵敏度(定义见“方法”); DSA, 删除-替换-加法; ENET, 弹性网络; EWAS, 全环境关联分析; EWAS-MLR, EWAS 多元线性回归分析; FDP, 假阳性率; GUESS, 图形单元进化随机搜索;  $n_B/k$ , 方法( $n_B$ )选择的变量数量除以真实预测变量的数量( $k$ ); sPLS, 稀疏偏最小二乘回归; TP, 真实预测变量。



[注]在每个被0~25个真实预测变量定义的情景下,各统计学方法运行100次,得到的灵敏度和FDP以均值表示。DSA,删除-替换-加法;ENET,弹性网络;EWAS,全环境关联分析;EWAS-MLR,EWAS多元回归线性分析;GUESS,图形单元进化随机搜索;sPLS,稀疏偏最小二乘回归。

图2 情景组1的灵敏度和假阳性率(FDP)

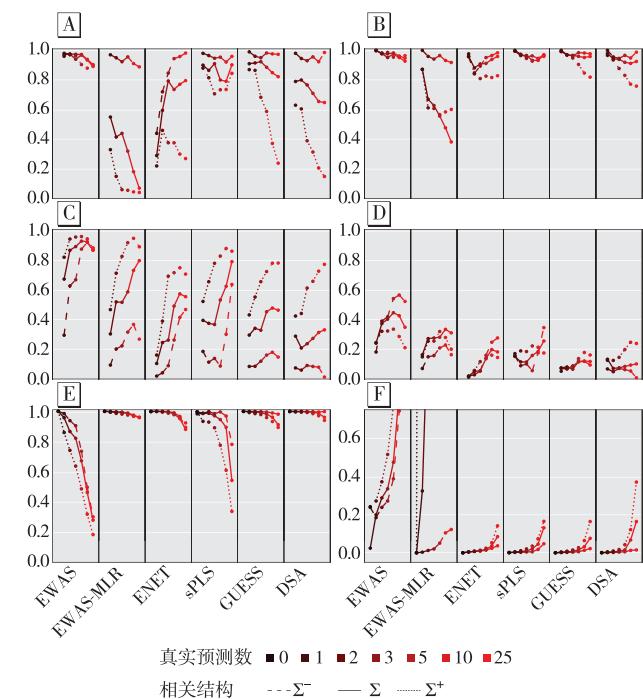
GUESS、sPLS、ENET和DSA方法的FDP都比EWAS或EWAS-MLR更低。计算所有情景下上述4种统计学方法的均值(第5百分位数,第95百分位数),灵敏度为78%(60%,91%),FDP为39%(21%,62%),特异性为96%(89%,100%),AltSens为95%(91%;99%),AltFDP为12%(5%,20%)。平均绝对偏差为0.03(0.00,0.11),仅考虑真实预测变量时为0.52(0.32,0.89)(图S2)。上述方法平均选择了1.79倍的真实预测变量(表1中的 $n_B/k$ )。DSA和GUESS在灵敏度和FDR之间的平衡(平均值分别为81%和34%)比sPLS和ENET(平均值分别为76%和44%)更好,其中DSA灵敏度稍高,GUESS的FDP更低(图2)。然而,没有一种统计学方法能在所有场景和所研究的指标上优于其他方法。

随着真实预测变量数量的增加,上述方法选择变量的性能通常降低:FDP和AltFDP在所有统计学方法中都增加( $k=1$ 和 $k=25$ 之间平均分别增加29%、9%);除EWAS-MLR和ENET( $k=1$ 和 $k=25$ 之间分别降低7%和4%)外,其他方法的灵敏度和AltSens均略有降低,平均绝对偏差增加(特别是基于EWAS的方法)。EWAS-MLR的灵敏度和AltSens大幅降低,而ENET的明显升高。然而在解释这些趋势时应注意,真实预测变量数量增加的同时,信噪比(真实模型的 $R^2$ )也增加,还伴随着真实预测变量高度相关的风险增加。

### 3.3 其他情景下的统计性能评估

选择真实暴露的情景组,其所有绝对成对相关性均<0.2(情景组2)或>0.5(情景组3),表明真实预测变量之间相关性水平越高,ENET、GUESS和DSA方法的灵敏度越低(EWAS-MLR方法幅度较小),而平均绝对偏差越高,主要表现在EWAS和DSA方法上(图S4)。ENET、sPLS和DSA方法对FDP有影响,但方向不一致。除了ENET方法AltFDP明显降低,特异性及灵敏度和FDP的代替指标都仅受到轻微影响。值得注意的是,选择高度成对相关的预测变量会导致在此模拟条件下误差的方差增加。

从具有较高(场景组4)或较低(场景组5)相关性水平的相关矩阵中生成暴露(图3)并未改变方法的可比性,但可对灵敏度、FDP和平均绝对偏差产生重要影响:暴露间的相关性水平越高,方法的统计性能越差。与情景组1相比,将相关性水平除以2,所有情景和统计学方法的灵敏度均>85%(除ENET  $k < 3$



[注]模型的性能以灵敏度(A)、替代灵敏度(AltSens,见“方法”)(B)、假阳性发生率(FDP)(C)、替代FDP(AltFDP,见“方法”)(D)、特异性(E)和绝对标准偏差(F)表示。实线上的数据点自有协变量矩阵 $\Sigma$ (情景组1)的多因素正态分布中产生;虚线上的数据点自协变量矩阵 $\Sigma^-$ (由 $\Sigma$ 矩阵的相关性除以2,情景组4)得到,点线由0~25个真实预测变量定义,运行100次得到的统计学指标在图中显示为均值(点)。DSA,删除-替换-加法;ENET,弹性网络;EWAS,全环境关联分析;EWAS-MLR,EWAS多元线性回归分析;GUESS,图形单元进化随机搜索;sPLS,稀疏偏最小二乘回归。

图3 依据暴露间相关性的数量分类后各统计学方法的性能

以外),与情景组1中的同一情景相比FDP平均降低23%。AltSens、FDP和特异性对暴露的总体相关性更不敏感,受到的影响也较为不一致。

正态分布暴露假设的偏差(情景组6)导致其结果与情景组1得到的结果类似,但EWAS-MLR方法通过自举数据得到了更好的结果,其他方法未能实现(图S5)。

真实预测变量的不同效应大小(从情景组7中[0.5, 1.5]的均匀分布得出)未改变方法的可比性且对统计学性能的影响有限:灵敏度和AltSens降低(与情景组1中同一场景相比平均-10%~-7%),特异性、平均绝对偏差(除了基于EWAS的方法)、FDP和AltFDP不受影响(图S6)。

### 3.4 变量选择的扩展方案

情景组1中,当由某种方法选择的变量数目增加,且该方法的变量与任一高于阈值 $\alpha$ 的变量相关联时,FDP大幅增加(基于EWAS的方法除外,因为其FDP已非常高),即使 $\alpha$ 高达0.8或0.9(表S2)。

## 4 讨论

为了从大量相关暴露中找到与连续健康结局有因果关系的暴露,本研究测试了若干种统计学方法的效力,主要依靠灵敏度和假阳性发生率来评估:模拟的特异性总是很高(至少有部分归因于研究假设,即在237个暴露变量中有不超过25个与结果相关),因而FDP成为了一项更具分辨力的标准。除了常规的灵敏度和FDP衡量指标,本研究还引入了替代指标以解释“与真实预测变量相关的假阳性可能实际上提供了可用于识别真实预测变量的信息”这一现象。

与EWAS相关的方法在所考察的情景下均表现不佳。无论使用何种程序校正多因素假设检验(Benjamini和Hochberg、Benjamini和Yekutieli、以及基于置换的FDR程序或Bonferroni校正),EWAS都混入大量的(假阳性)协变量(情景组1中平均FDP为86%)。高FDP的原因主要是FDP程序假设统计学指标(此处指 $p$ 值)是无偏倚的,而在本研究的模拟中,由于将回归模型独立地拟合在相关暴露上,存在混杂的可能性很高。然而与其他方法相比,EWAS在估计真实预测变量的系数值上表现最佳。当EWAS后加一个多元线性回归步骤(EWAS-MLR)时,能识别小部分真实预测变量(情景组1的平均灵敏度为33%)。但上述两种方法仍优于不应用校正的多重比较,表现为

各暴露与不同文献中连续考虑的结局之间的相关性。这两种方法的AltFDP仍然相对较高(情景组1中平均32%),表明在所考察的情景中,许多被上述方法错误选择的变量与真实预测变量之间的相关性不强。

使用ENET、sPLS、GUESS和DSA方法能筛选出大部分真实预测变量(情景组1的平均灵敏度为78%),但相当大比例的暴露被错误地认为与结局相关(情景组1的平均FDP为39%)。被上述四种统计学方法错误选择的暴露,平均至少与一个真实预测变量高度相关(情景组1的平均AltFDP为12%)。同样地,当一个真实预测变量未被上述方法选择时,很可能另外选择了一个与其高度相关的协变量(情景组1的平均AltSens为95%)。在所有考察的情景和条件下,没有一种多因素统计方法明显优于其他方法。总体上,DSA和GUESS在灵敏度和FDP之间展现出更好的平衡,其中DSA的灵敏度更高,GUESS的FDP更低。背离暴露正态分布或真实预测变量效应大小均匀的假设不会改变方法的可比性,但是GUESS和DSA受真实预测变量间高水平相关的影响最大(情景组3),而EWAS和sPLS则对其较不敏感。选择上述方法时其他因素也很重要,例如易于使用、影响混杂因素的能力以及适应不同研究设计(例如纵向研究)或非线性的暴露-反应关系(例如使用样条函数)。

有观点认为,选择与真实预测变量高度相关的变量不应被认为有本质上的错误(Frommlet等,2012),本研究对FDP和灵敏度的替代定义实际上是在此逻辑下制定的。从四种多因素统计学方法修订后较高的检验标准来看,大多数真实预测变量很可能属于一组与所选变量高度相关的暴露。因此,考虑“真实自变量”及其相关协变量可能是识别真实预测变量的一种方式。考虑这类变量选择扩展方案时应当注意以下几点:a)遗传学研究中,可利用基因组的结构识别已知和未知的相关多态性;这种技术可能不适用于暴露组研究,因为暴露之间的相关性可能源于多种机制(饮食、社会经济地位等),不能保证选择相关的变量就能提供关于真实预测变量和结局之间因果机制的有用信息。因此,区别真实预测变量和与真实预测变量相关的变量是一大挑战;b)降低选择的阈值(通过纳入所有与所选预测变量相关的变量)可能导致FDP在通常定义下升高,从而抵消了其有利的方面(即灵敏度增加)。这种效应在本研究方案的结果中已有显示,即选择与自变量相关性水平>0.8的变量时FDP明

显增加(表S2)。鉴于此,需要着重强调FDP的替代指标(AltFDP)不是由AltSens诱导的变量选择方法产生的FDP,而与所选变量高度相关的预测变量同样会被选择。相反,FDP是由原始选择方案产生的,但将相关变量计为“真实预测变量”,其权重与他们和真实预测变量的相关性成比例。

本研究的模拟以Sun等(2013)的研究为基础,就暴露的数量及其相关结构设定了更真实的暴露组情况。结果显示,生成暴露的相关结构极大地影响统计学方法的性能(图3),说明Sun等及其他相关结构固定的模拟研究结果不能直接外推到暴露组情况中。

本研究建立在若干个模型假设的基础上,在讨论结果的外推性时必须考虑这些假设。首先,假设任意其他协变量都不对产生健康结局的协变量具有效应修饰作用(违背加和性),这实际上是不可能的。合并相互作用项将大大增加模型空间的大小(例如本研究中的27996个一级交互作用),同时需要扩展统计学方法以检验交互作用,使用来自研究中所考察的全部技术(例如,Li和Zhang,2010)。在模拟中取消对二分类效应大小的限制和引入的可变效应大小不会改变FDP,仅人为降低了检验弱效应的统计学性能(灵敏度平均降低10%)。上述结果可解释为天花板效应,换而言之,对高于平均效应的暴露,已经较高的灵敏度不会再提高,正如对低于平均效应的暴露,灵敏度不会再降低。总体而言,所有方法的灵敏度都下降,无法进一步区分被考察的统计学方法。重要的是,本研究未考虑暴露协变量的测量误差或错分,尽管它们可能对统计学性能或偏移产生很大影响,特别是经典误差(de Klerk等,1989; Perrier等,2016; Rappaport等,1995)。因此,在实际情况中,方法的统计性能可能减弱,但此前没有证据证明本研究考察的统计学方法会受到上述问题的不同影响。本研究进一步假设暴露是正态分布的,背离假设也不会改变方法的统计性能。最后,与Sun等(2003)的结果类似,本研究

使用的一组有限的统计学方法都来自线性回归框架。轮廓回归、聚类分析或其他机械学习方法等替代方法可作为补充,但不能与本研究中基于回归的方法直接做比较。

## 5 结论

本研究根据一个真实的暴露组结构,筛选出大量相关的暴露,其中只有少数与连续效应直接相关。结果表明,所考察的多因素方法在研究暴露组方面优于单因素方法:尽管未实现低FDP,但上述方法的统计学性能令人满意,并在不同程度上保持了灵敏度和FDP之间的平衡。就本研究采用的指标来看,DSA和GUESS的统计学性能更好,但并非在所有情景或条件下均如此,且在实例分析中,方法选择也受到计算复杂性和灵活性考虑的影响,例如容纳混杂的能力。统计学方法的性能很大程度上受暴露组协变量之间相关性的影响,提示暴露组研究中的一个固有问题,即研究的统计学方法不能有效区分真实预测变量和协变量。

**编者注:**作者指出,为更好地对情景组1的结果进行总结,经修订后表1中的各统计学指标以七个情景的均值(最小值,最大值)表示,修订前为情景组1中七个情景在所有模拟条件下(每个情景运行100次)的均值(第5百分位数,第95百分位数)。

(续完)

翻译:窦冠坤;审校:金泰廙

## 参考文献(略)

本文原文刊登于EHP杂志,需要者务必引用英文原文,详见Agier L, Portengen L, Chadeau-Hyam M, et al. A systematic comparison of linear regression-based statistical methods to assess exposome-health associations. Environ Health Perspect 124( 12 ):1848-1856.

本文原文及参考文献请浏览<http://dx.doi.org/10.1289/EHP172>

(编辑:汪源;校对:陶黎纳)