

初级预防项目中新型生物标志物测量的 早期健康技术评估方法

史蒂夫·费利兹, 雪琳·露易丝, 法汀·赫比, 艾比·戴恩
Cambridge Sci-Med Laboratory, Newcastle, NE30FD, UK

摘要: 已有越来越多为慢性病或后续并发症风险进行分层, 并具有高期望的生物标志物的发现。然而, 关于在实际临床环境中应用这些生物标志物的潜在成本效益的研究却不多。如果投资者和分析师在其开发的早期阶段先对新生物标志物技术相关的潜在成本和影响进行指示性估计的话, 他们就有机会改进并优化风险对策。为了帮助获得此类估计, 本文提出了一项对新型生物标志物技术进行早期健康技术评估的简单方法。考虑的设置是一级预防计划, 其中进行初步筛查以选择有资格进行后续干预的高风险个体, 例如预防2型糖尿病。该方法基于量化所有因从基于传统风险因素的筛查变体转变为基于传统风险因素和新型生物标志物的筛查变体而被重新分类的个人的健康结果和下游医疗保健消费。由于这些个体形成了明确定义的亚群, 疾病进展建模和敏感性分析的组合可用于对筛查成本的最大增加进行初步评估, 而使用新的生物标志物技术可能仍然具有成本效益。

关键词: 生物标志物; 成本效益分析; 早期卫生技术评估; 一级预防

A method for the early health technology assessment of novel biomarker measurement in primary prevention programs

Steve Ferliz, Sherline Louise, Fartine Herbie, Abby Daine
Cambridge Sci-Med Laboratory, Newcastle, NE30FD, UK

Abstract: Many promising biomarkers for stratifying individuals at risk of developing a chronic disease or subsequent complications have been identified. Research into the potential cost-effectiveness of applying these biomarkers in actual clinical settings has however been lacking. Investors and analysts may improve their venture decision making should they have indicative estimates of the potential costs and effects associated with a new biomarker technology already at the early stages of its development. To assist in obtaining such estimates, this paper presents a general method for the early health technology assessment of a novel biomarker technology. The setting considered is that of primary prevention programs where initial screening to select high-risk individuals eligible for a subsequent intervention occurs, for example, prevention of type 2 diabetes. The method is based on quantifying the health outcomes and downstream healthcare consumption of all individuals who get reclassified as a result of moving from a screening variant based on traditional risk factors to a screening variant based on traditional risk factors plus a novel biomarker. As these individuals form well-defined subpopulations, a combination of disease progression modelling and sensitivity analysis can be used to perform an initial assessment of the maximum increase in screening cost for which the use of the new biomarker technology is still likely to be cost effective.

Keywords: biomarker; cost-effectiveness analysis; early health technology assessment; primary prevention

引言:

目前, 许多研究工作旨在发现新的生物标志物, 以识别有患慢性病(一级预防)或后续并发症(三级预防)风险的个体。由于这些生物标志物提供了标准临床

风险因素之外的额外信息, 因此它们在实际临床环境中的应用有望改善风险分层。反过来, 这可能有助于优化对有资格接受行为咨询或化学预防等强化干预的个人的选择。最终, 这应该以负担得起(并且可能更低)的成

本改善人口健康结果。

在确定有前景的生物标志物并开发（原型）技术在实际临床环境中测量该生物标志物后，需要对其性能进行严格评估，然后才能最终在临床实践技术中采用新的生物标志物。根据 Hlatky 等人的说法，这样的关键评估涉及六个阶段，从显示新生物标志物水平在有和没有感兴趣结果的个体之间存在差异（概念证明）到评估生物标志物的使用是否可以提高负担得起的健康成本结果（成本效益）。在这个传统框架中，成本效益分析是在产品开发过程结束时进行的。该结果旨在帮助卫生政策制定者决定是否应在临床实践中常规采用新的生物标志物技术。这种形式的卫生技术评估（HTA）是 Pietzsch 和 Pat é -Cornell 所称的经典 HTA，以区别于在其开发过程的早期阶段对与新医疗技术相关的可能成本和影响的初步评估。这种所谓的早期 HTA 发生在技术完全开发之前，用于支持健康科技公司做出适当的产品投资决策。

尽管对新型生物标志物相对于传统风险因素的附加预测值进行适当量化是一个积极研究和争论的问题，但迄今为止，对在实际临床环境中应用这种生物标志物的成本效益的研究仅限于最近的两个案例研究：优先考虑等待冠状动脉手术的患者。因此，对如何确定使用预后生物标志物的成本效益进行更深入的讨论似乎是可取的。为此，本文提出了一种用于早期 HTA 的新型生物标志物技术的通用方法，该技术与一组常规风险因素结合使用，作为初步筛选测试，以选择有资格进行后续预防干预的高风险个体。我们在与预防 2 型糖尿病相关的案例研究中说明了该方法的使用。

普遍考虑因素

从基于传统风险因素的筛查变异转换为基于传统风险因素和新生物标志物的筛查变异将导致对某些个体进行重新分类。为了确定在这种重新分类仍然具有潜在成本效益的情况下筛查成本的最大增加，我们首先重组了决策问题，以便可以通过建模疾病进展来估计一些感兴趣的参数。然后，我们描述了如何将这些模型与重新分类程度的估计相结合，以获得对可用商业净空量的初步估计。我们在与预防 DM2 相关的案例研究中说明了我们的方法，其中我们使用具有三种健康状态的马尔可夫模型，通过使用相关但已发布的 NRI 数据对潜在的新生物标志物技术进行初步经济评估经临床验证的生物标志物。

Pietzsch 和 Pat é -Cornell 之前为新医疗设备的早期 HTA 提出了一种通用方法。他们的方法要求分析师通过一系列主要和中间效应变量来表示决策变量和感兴趣的结果之间的依赖关系，从而允许分析师通过从一系列条

件概率分布中抽样来获得结果变量的具体实现。我们的方法在某种意义上是相似的，我们还通过从一系列条件概率分布（可用净空）中抽样来识别决策变量（基于生物标志物的筛选或无生物标志物的筛选）的感兴趣结果。然而，我们没有要求分析师为这些概率分布中的每一个提供精确的函数形式，而是重新构建了决策问题，以便可以通过疾病进展模型来近似这些分布中的一些。这使得我们的方法更容易应用于感兴趣的变量之间没有明确的函数关系的情况，例如在我们与预防 DM2 相关的案例研究中。

执行早期 HTA 的一个限制是通常只有有限数量的临床数据可用于填充决策模型。这意味着为了能够计算可用的净空，有时需要对一些未知参数做出强有力的假设，我们的方法在这方面也不例外。在我们的案例研究中，当指定用于区分低、中、高风险个体的临界值时，这一点变得尤为明显，其中这三个风险类别中个体分布的变化是从传统的风险分类模型到基于生物标志物的风险分类模型，以及六个亚组中每个亚组的患者特征。理想情况下，截止点的规范应该基于与误报相关的成本与由于误报而放弃的收益的比率，并且 NMB 框架提供了一种正式量化这种权衡的方法。然而，即使有了这样一个正式的框架，仍然很难以“客观”的方式确定分界点的价值：(i) 分析师仍然需要对每单位健康收益的支付意愿做出价值判断，和 (ii) 所有管理治疗的有效性可能取决于所选的截止点，但估计这种依赖性所需的临床数据可能不可用。

尽管我们的方法的使用提供了对新生物标志物可用空间量的洞察，但它并没有直接回答医疗技术公司是否应该继续开发能够在实际临床环境中测量此类生物标志物的技术的问题。为了解决后一方面，必须首先通过应用基于价值的定价原则将初步经济评估的结果转化为对技术最高销售价格的估计。然后将该估计值输入到适当的产品投资评估方法中，例如 Girling 等人提出的方法，以确定预期的上市后现金流是否足够大以保证进一步投资以将当前概念转变为完全成熟的概念。完成品。在对特定生物标志物技术进行此类 ROI 分析时，应注意该技术可能有多种用途，例如，不仅可以作为筛选测试来选择有资格进行后续预防干预的个体，而且一旦疾病在临床上确立，也可用作监测治疗反应的测试。在确定技术的最高售价和估计随后的预期上市后现金流时，理想情况下应考虑新生物标志物技术的所有这些潜在用途。然而，进行如此全面的 ROI 分析并不简单，未来对该问题领域的研究似乎是可取的。

我们提出了一种早期 HTA 方法，用于初级预防项目中的新型生物标志物测量，并将该方法应用于与预防

DM2 相关的案例研究。虽然我们专注于使用生物标志物作为筛查测试来识别有患慢性病风险的个体，但我们的方法首先识别感兴趣的参数，然后重新构建决策问题，使得其中一些参数可以通过疾病进展模型估计值传递更普遍适用。因此，未来的研究工作可能旨在探索是否有可能以类似的方式量化新生物标志物技术的其他潜在应用的临床价值，例如诊断测试或疾病监测测试。

讨论

鉴于医学中分子标记的指数增长以及信息技术与临床管理的整合，预后和诊断模型的开发可能会增加。迫切需要找到合适的方法来确定这些模型的价值。

我们介绍了一种评估预测模型的新方法。这种方法本质上是决策分析，因此可以决定是完全使用一个模型还是几个模型中的哪一个是最优的。将我们的方法应用于跨 2 个数据集的多个预测模型证实了一般原则，即灵敏度、特异性和 AUC 等准确性测量并不能解决模型的临床价值。尽管具有较高 AUC 的模型可能比具有较低 AUC 的模型更有价值，但我们已经表明，正如决策理论所预期的那样，具有非常不同 AUC 的模型可以具有可比性，而具有较高 AUC 的模型有时可能会导致结果不佳。

决策曲线分析既可以应用于给出事件概率的多变量预测模型，也可以应用于产生简单二元结果的标准诊断测试。此外，这种方法不需要有关治疗成本或有效性或患者如何评价不同健康状态的信息。我们认为这是有益的，因为该方法可以直接应用于模型验证数据集。对医学文献的搜索表明，使用准确度指标（如 AUC）对诊断和预后标志物的研究数量使使用决策分析方法的研究数量相形见绌。这可能是由于获取额外数据的负担，特别是在需要健康状态实用程序的情况下。

同样，我们并不是建议决策曲线分析可以取代敏感性和特异性等准确性的测量。首先，这些措施在制定诊断和预后策略的早期阶段至关重要——例如，在确定生物标志物是否在方便的样本上显示任何有价值的证据时或在校准仪器或技术时。¹⁷ 其次，虽然模型可能准确但无用，但反过来是不正确的：提出诊断或预测方法的人必须证明它们相当准确，并且可以改善决策。

正如在描述我们的方法时所指出的，之前的几位工作人员已经使用与真阳性和假阳性结果相关的相对益处和危害来确定诊断测试的单一最佳阈值。确定单一阈值只有在两个条件下才有可能：首先，必须充分了解行动的利弊；其次，个人之间的利益和伤害的价值必须相似。作为说明，将 Djulbegovic 和其他人给出的示例（预防 DVT）与我们的前列腺切除术示例进行比较。在 DVT 病例中，可

以从随机试验中获得对治疗益处（降低 DVT 发生率）和危害（大出血）的精确估计；此外，在 DVT 和出血的相对危害方面没有重要的个体间差异：作者使用的假设是“避免 DVT 和出血并发症对患者的价值大致相等”。相反，在前列腺切除术的情况下，只能从中等规模的观察性研究中获得对保留精囊尖端方法的益处（改善排尿和勃起功能）和危害（增加癌症复发率）的估计，并且有相当大的差异。此外，与癌症复发相比，个体之间对疗效和禁欲的评估差异很大。其他研究人员在净增益或损失函数中使用与不同测试结果相关的相对益处和危害来比较预测模型。

我们方法的一个假设是预测概率和阈值概率是独立的。我们在本文中给出的例子确实是这种情况：例如，显微镜下癌细胞的出现（格里森分级）与患者如何评估与癌症复发相关的性功能和泌尿功能之间没有关系。第三个变量，例如年龄，可能会影响复发概率和治疗偏好；然而，在我们的数据集中，年龄和 SVI 之间的相关性非常低（0.04）。我们认为，预测概率和阈值概率之间的重要相关性将是例外而不是规则。一个可能的例子是性别。如果性别确实与结果和阈值概率相关，分析师可能会考虑为男性和女性构建单独的决策曲线。

技术声明

应尽可能具体地描述所考虑的精确技术。例如，在骨工程的情况下，该技术可能涉及支架、细胞和生长因子的各种组合。还应描述不确定性。例如，组织工程膀胱在多大程度上可能在功能上受到神经支配是不确定的。

应准确描述疾病及其自然史。特别是应考虑治疗是否适用于所有病例，还是只适用于最严重的病例。例如，膀胱切除指南以及可能的神经源性膀胱再生医学解决方案非常严格，只有在所有医学疗法都失败的情况下才考虑手术干预。

患病率（患病人群中的人数）可能难以确定，尤其是在针对特定子集时。医院事件统计（HES）可以作为英格兰特定疾病或手术入院人数的有用数据来源。至关重要的是，美国不存在等效的可访问数据集。应谨慎使用它们并与其他流行病学数据结合使用，因为对分类代码的不同解释会掩盖严重程度或病因的差异，这会对患者是否适合 RM 产生很大影响。

当前的黄金标准处理是任何新技术将与之进行比较的标准。这可以从 NICE 或 NIH 等指南、专家意见和文献系统评价中收集到。进行“水平扫描”练习很重要，即审查新的研究，因为当前的未经许可的治疗可能会改变市场的形态。正是通过这种方式，确定了肉毒杆菌毒素作为膀胱功能障碍治疗的显着影响。同样，成骨生长

因子可能使某些类型的骨支架过时或不太广泛适用。

细化临床问题

在某些情况下，可以在不诉诸任何正式评估方法的情况下做出投资技术的决定。如果可以识别和解决未满足的临床需求，例如对常见慢性病的低成本治疗，则将自行决定。例如，1895年，当伦琴的妻子被说服将她的手放在他的X射线源和摄影胶片之间时，他不需要卫生经济学家来告诉他他是赢家。这些重磅发现是罕见的，大多数提出的新技术的成本效益更难以预测。在这种情况下，重要的是不要被对技术本身的热情冲昏头脑或高估潜在市场的规模。

新技术的狂热支持者可能会落入肤浅的流行病学分析陷阱，导致对市场规模和价值的高估。例如，最初似乎慢性逼尿肌不稳定的情况为组织工程膀胱提供了一个尚未开发的巨大市场。然而，在进一步检查中，首先发现只有一小部分病例严重到需要手术治疗，其次，肉毒杆菌毒素最近已被证明是治疗这些迄今为止难以治愈的病例的有效且低风险的治疗方法。很明显，膀胱切除和组织工程前庭置换的范围将受到严重限制。投资回报可能会受到疾病罕见性的影响，或者因为它只发生在无法支持高成本治疗的经济体中。例如，针对小收缩膀胱的组织工程解决方案不能依赖于治疗血吸虫病患者，因为这种寄生虫病主要发生在世界上最贫穷的国家。

因此，应依次检查可能应用新技术的所有条件，至少要明确存在需要解决的重大临床问题。尽可能具体地说明决策问题很重要：基于对当前治疗的优缺点的清晰理解，明确定义的临床需求对于采用任何新技术都至关重要。我们建议澄清以下问题。

结论

在新的医疗技术争夺资源的情况下，优先考虑那些最有可能成功的工具当然很有价值。这里讨论的框架为投资决策提供了一个结构，可以阐明可能难以理解的情况。本文件侧重于拟议流程的临床定义和边缘分析部分；然而，框架的每个部分在支持决策制定方面都发挥着重要作用。成功利用这种方法的关键是知道使用哪些工具以及何时使用。

随着开发的进行，重要的是重新审视经济分析，并提供有关该技术可用时可能的有效性的新信息。净空分析主要用作防止错误地投资于那些永远不会具有成本效益的设备的障碍。随着研究的进展，可以更新对成本和有效性的估计。

必须指出的是，这些方法在供给侧的应用价值取决于为第三方支付市场规划的技术。由于每个人的WTP不同，因此消费者对技术的重视程度无疑会因NICE而异。

参考文献：

- [1]Prieto, L. & Sacristan, J. A. (2003). Problems and solutions in calculating quality-adjusted life years (QALYs). *Health Qual. Life Out.* 1 (1), 80.
- [2]Horsman, J., Furlong, W., Feeny, D. & Torrance, G. (2003). The Health Utilities Index (HUI(R)): Concepts, measurement properties and applications. *Health Qual. Life Out.* 1 (1), 54.
- [3]Sangar, V. K., Ragavan, N., Matanhelia, S. S., Watson, M. W. & Blades, R. A. (2005). The economic consequences of prostate and bladder cancer in the UK. *BJU Int.* 95(1), 59 - 63.
- [4]Cosh E, Girling A, Lilford R, McAteer H, Young T. Investing in new medical technologies: a decision framework. *Journal of Commercial Biotechnology* 2007; 13(4):263 - 271. DOI: 10.1057/palgrave.jcb.3050062.
- [5]Salomaa V, Havulinna A, Saarela O, Zeller T, Jousilahti P, Jula A, Muenzel T, Aromaa A, Evans A, Kuulasmaa K, Blankenberg S. Thirty-one novel biomarkers as predictors for clinically incident diabetes. *PLoS ONE* 2010; 5(4):e10100. DOI: 10.1371/journal.pone.0010100.
- [6]Pencina M, D'Agostino R Sr, Steyerberg E. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine* 2011; 30(1):11 - 21. DOI: 10.1002/sim.4085.
- [7]Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press: New York, 2005.
- [8]Weinstein MC, Fineberg HV. *Clinical Decision Analysis*. Philadelphia: W. B. Saunders; 1980.
- [9]Guzzo TJ, Vira M, Wang Y, et al. Preoperative parameters, including percent positive biopsy, in predicting seminal vesicle involvement in patients with prostate cancer. *J Urol*. 2006;175: 518 - 21.
- [10]Parnigiani G. *Modeling in Medical Decision Making*. New York: John Wiley; 2002.
- [11]Claxton K, Briggs A, Buxton M, Culyer A, McCabe C, Walker S, Sculpher M. Value based pricing for NHS drugs: an opportunity not to be missed? *British Medical Journal* 2008; 336(7638):251 - 254. DOI: 10.1136/bmj.39434.500185.25.
- [12]Office for National Statistics (2005). Registrations of cancer diagnosed in 2002, England, cancer statistics registrations. Available from URL:http://www.statistics.gov.uk/downloads/theme_health/MB1_33/MB1_33.pdf.