白皮书

蛋白生物标志物研究的设计策略

引言

以蛋白生物标志物为基础的研究在推动精准医疗的发展上具有巨大潜力，但需要仔细考虑研究设计和数据统计等以最大化其影响。

对合理研究设计的要求包括一个明确定义的研究目标、足够的样本量、干扰因素的控制，以及有偏差且恰当的统计分析。而其中最重要的一点是应用标准化的样本处理流程以确保同等的样本质量。在Olink *白皮书《蛋白质生物标志物研究的分析前影响因素》*[1]中提供了关于样品处理和操作的建议。

在研究设计时应充分考虑该研究涉及到的所有步骤，从最初的规划，到样本的收集和检测，直至最终报告。如果一项研究设计得当，那么任何可能使测试过程的结果扭曲或偏差的因素都可以被最小化。本篇白皮书描述了当你计划未来的研究时，在研究方案的设计中需要考虑的重要因素。

计划阶段

研究的计划阶段非常重要。为了使研究结果和结论有信心， 项目的研究目标和其它在成功的结果中需要考虑的因素都应在项目的最开始被指出。例如，如果不能信心满满地得出强有力的结论，那么这种力不从心的研究将可能导致有限资源被浪费。

术语

在群体比较研究中，研究的效能一般为被检测到蛋白质水平的真实差异与其在假设检验中具有统计学意义的概率。效能分析可以用来计算获得有力结果所需要的最小样本量。通常建议使用尽可能高的效能，但在实践中，基于效应量等参数假设，研究设计往往采用80%的效能。

所有研究都有偏差，而研究者应始终正视和避免偏差。因此实验设计必须以统计学角度作全面的考虑。我们建议跨学科团队一起合作设计研究方案。

研究的效能

为了将实验目的转化为统计或分析计划，了解如何为研究获得准确的效能计算是非常重要的。不同的研究问题需要不同的统计方法，而这些方法又需要不同的样本量来检测到显著变化。统计方法的使用可能对效能分析有显著影响。如果不了解最适合特定实验的统计学工具，就不可能 实现恰当的效能计算。在许多最常见的统计检验中，效能计算可以通过一个简单的方程式解决， 参见Dell *et al.*[2]， 但对于更复杂的分析来说，统计模拟可能是非常必要的。

效能分析

在理想情况下，效能分析往往在样本运行和收集数据前进行。 在正式研究实验前的效能分析通常用来确定要获得足够有力的结果所需要的合适样本量。效能分析也可以在研究完成之后进行，这种回顾性的计算在从小型预试验扩大到大型研究时可能是有用的。

效能计算也可以用来确定最小可测量的效应量，这意味着在该最小样本量下的研究可能被检测得到两组之间的差异量。

样本量计算

在研究的规划阶段进行样本量计算的目的是确保结果能够提供有用信息。如果样本量太小，产生有意义结果的可能性将会很低。而另一方面，样本量太大将导致不必要的成本浪费和时间浪费，也可能被认为是不道德的。

如果一项结果非常罕见并鲜少发生，则必须通过增加样本量来检测潜在差异。由于效能在很大程度上由研究中最小的群体决定，对稀有组别采用过量样本也是明智之举。因此推荐在不同组别之间设计平衡研究。

增加样本量就像是在增加一幅图片的分辨率。如果只有几个样本，图片将会非常模糊，我们只能够区分那些差异最显著的数据。 然而当样本量非常大时，整张图片就会变得足够清晰，甚至使我们能够确认数据之间非常细小的差异。

变量

效能与离散程度（标准差）、效应量、样本量和显著性水平直接相关。下文中将对这些变量进行描述。一般而言，如果在五个变量中有四个变量保持不变，就可以计算或估计剩下的那个变量。在现实中，我们通常会探索多个级别的常量变量，以了解目标变量（如效能）是如何受到影响的。

重要信息！

在研究之前所进行的全部效能分析都建立在变量的*假设*基础上，只能作为研究设计时的指导。这些假设可能来自于领域知识、历史数据、文献检索或预实验。

离散性

数据的离散性通常用标准偏差衡量，其描述了观察到数据点的离散程度，在进行效能计算时必须进行估计。如果一个变量的标准偏差太大，那么与具有更小标准偏差的变量相比，想要获得一定的效能就需要更大样本量。

一项研究可能存在多个可变来源，因此离散性可能会难以估计。如果没有很好的估算方法，那么在低水平和高水平的可变性程度下进行效能分析，看看它对计算出的样本量或效能有何影响是非常值得的。

效应量

效应量是可以标准化的，它包含测量的可变性信息。

效应量标准化的一个常用方法是用观察到的平均差异除以估计的标准偏差，该值被称作Cohen's *d*。如果实验组与对照组之间的数据没有重叠，则存在实质性差异，效应量高。但如果两组之间的重叠大于其差异，效应量则不显著。

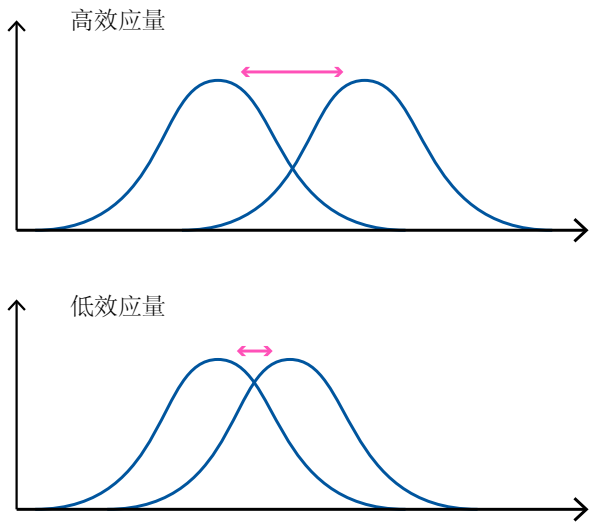

图1. 高效应量将两种分布拉开，更容易识别差异。

效应量的标准化参数，Cohen's *d*，相当于标准正态分布中的Z-score。如0.8的效应量意味着实验组的平均得分比对照组的平均得分高出0.8个标准差。检测规模越大效能越高。因此，如果效应量很小，则需要更多的样本才有检测能力。

样本量

样本量是研究中每组样本的数量。它通常是我们想要估算的一个变量，通常也是研究人员控制研究效能的唯一的因素。然而，在不可能增加样本量的情况下，可以使用效能分析来确定最小的效应量，此时提出的研究应拥有合理的高效能。而增加样本量往往能增加研究的效能。

显著水平

显著水平（通常用α表示）是假阳性结果的概率。在进行统计学检验时，显著性水平通常设置为0.05。然而， Olink的每块 panel都含有92重检测，每个检测通常使用一次统计学检验。 这意味着在多重检验时，应该相应地调整显著性水平。一种常用方法是用显著性水平除以测试的数量，这被称为Bonferroni校正。在对92个检测修正后的校正水准约为0.000544。

每个数据都有一个相关的概率值，称为p值。它被定义为在给定的抽样分布下，当不存在真正的差异时，偶然观察到差异的概率。

如果经必要校准后的p值低于显著性水平（$p < 0.05$），则结果具有统计上的显著差异；如果p值比显著性水平高（$p>0.05$），结果在统计学上无显著差异。增加显著水平将会提升效能，但也会增加在无差异时检测到差异的概率。因此，当增加检测数量（运行更多的panel）时需要运行更多样本来保持同等的效能水平。

在我们讨论的四个变量中（离散程度、效应量、样本量和显著性水平 ），目前效应量对效能的影响最大，但所有的变量都应被考虑在内。

这并不是全部的情况

统计显著性检验的一个主要缺点是对研究设计的盲目性。它们无法说明全部的情况。举例来说，如果一项研究治疗组只有70岁以上的男性，而对照组只有50岁以下的女性，那么将无法得知组间差异来源于年龄，性别还是正在研究的治疗。在这项假设研究中，研究人员仍然可以计算出p值，但结果不会有用处。

一些需要在规划阶段考虑的可能的干扰因素：如年龄、性别、其他基础疾病，或样品处理的差异，即所谓的分析前差异。

考虑不同临床研究类型

公认的临床研究有几种不同的类型，每一种都有各自的优势和挑战。研究范围可以从基础疾病与对照比较研究（病例对照研究）到大队列研究，这些研究收集了丰富的临床数据，并可能包括一系临床终点。

研究可以是前瞻性的 （在研究开始之前计划好所有细节，如测量参数或要进行的临床干预措施）或回顾性的 （材料来自于已完成的研究，在事后进行调查或重新询问）。

如果在研究过程中从同一个体的多个时间节点提取样本，则被称为纵向研究。我们将会考虑到与不同类型的临床研究相关的研究设计的中某些方面。

队列研究中的发现和验证

Olink的panel是对大量蛋白进行探索性筛选的理想选择，研究人员在其中寻找被测蛋白间的交互模式或相对差异。

这种广泛筛选方法的一个例子来自于Bryan等人进行的一项研究，在研究中运行了五个不同的Olink panel。他们发现了与尿路上皮膀胱癌显著相关的全新蛋白质，和可能的预后分级候选标志物[3]。

在临床环境中，病例通常有症状，且经历过各种诊断程序。对照组可能包含患有其它疾病的患者，或由他们组成， 病例组和对照组之间的样本收集和处理程序可能存在差异。根据样本收集的人口统计学特征，在特定队列中也可能产生遗传偏差。以上所有因素都可能对研究结果产生潜在影响。

在基于蛋白标志物的队列研究中，结论的置信水平可以通过在第二个独立的队列或样本中验证结果来获得大幅提升。在众多已发表的此类多队列研究方法中，一个典型的研究来自于Tromp等人，他们发现了一些可以区分两种主要心力衰竭类型的蛋白标志物。在该方法中，蛋白标志物的发现首先来自于一个~1500个样本的队列，然后在一个含有850个样本的独立队列中得到验证[4]。

纵向研究中不同批次与不同序列间的差异性

在纵向研究中，每隔一段时间评估个体或群体间的相关差异，同时持续或反复监测风险因素与健康结果。

一个关键的考虑因素是不同批次间的重现性。试剂批次的改变可能在不同批次产品的运行间产生偏差。同样的，即使使用同一批次的试剂，如果实验室条件发生波动，在纵向研究节点之间也可能引入偏差。因此通常建议在纵向研究的每个时间点包含桥接样本。为这些样品测量的蛋白质水平可以用作批次之间的共通用参考，并用于规一化和减小任何潜在偏差。

术语

桥接样本是在不同批次间样本板运行的样本。桥接样本应代表关于样本和对照的数据集。这些样本用于参考样本归一化。Olink建议每个案例中至少使用8个不同的桥接样本。

要确保有足够的样本供应以保障研究间的桥接。建议您与Olink的代表进行完整的研究讨论。以保证实验中包含合适数量的试剂盒和桥接样本。

在新试剂盒的生产中，Olink采用QC流程来限制批次间的变化，并确保一致的性能。这一点很重要，因为新的抗体批次可能会导致生成的数据发生变化。为了解决这个潜在的风险，Olink引入了完整的质量控制程序和严格的验收标准。

然而，当试剂盒被用于纵向研究时，在统计分析中必须考虑和调整多个批次之间的潜在信号差异。

Olink将不断完善产品和服务。有关panel组成和更新的最新信息，请联系您当地的Olink代表或support@olink.com。

样本的注意事项

在采用Olink panel进行任何研究之前，有几件事需要注意。

样本和对照

在研究中，病例组和对照组应该使用相同的样本基质，以便能比较组间的数据。不同基质之间的数据无法直接比较，但可以确定部分内容，例如在血浆中发现的生物标志物是否也可以在脑脊液中被测定。需要设计对照组的纳入标准以适应研究问题。对照组通常由健康个体样本组成，但这并不总是最好的选择。例如在一项评估疾病二级预防的潜在生物标志物的研究中，健康对照就不是一个很好的选择。

正如Yeh等人指出的那样， 对照组的设计可能会很棘手。他们评估了匹配的兄弟姐妹病例和对照研究设计是否会在发现乳腺癌重要的早期诊断生物标志物方面产生更多的统计效能。 他们假设姐妹可以作为良好匹配的对照，因为她们自然地受到种族、民族和很大比例的遗传背景的控制。然而，来自生物学姐妹的样本在整体上并没有比其它个体样本具有更高相似度，也没有很好的与其它样本区分[5]。

重要的是要记住，在解释数据时，关于样本处理的适当文档可以提供有价值的输入。请阅读*白皮书《蛋白质生物标志物研究的分析前影响因素》* 获得更多信息[1]。

质量控制

应选择合适的质量控制。许多其它的多重免疫分析方法在试验后都需要进行验证步骤。而为了避免这种情况，Olink开发了一个内置的QC系统，可以采用内部控制确保其多重标志物panel的检测质量。该系统可以完全控制分析和样本的技术性能。

除了Olink提供的对照外，所有板上还应该包括一个混合血浆样本或另一个定制的混合生物基质，以便进一步质控。 附加QC的一个案例是评估板间或板间的潜在差异，如检测板间CV或板内CV。阅读Olink*白皮书《数据的归一化与标准化》* 获得更多关于QC系统的信息[6]。

考虑最佳稀释度

每个Olink panel都优化了血浆和血清中92个生物标志物测定的预期生理和病理范围。 我们的一部分panel被设计为稀释运行（从1:10到1:2025），使可测量范围符合生理范围。这些稀释是测定的一部分，不需要预先稀释。

稀释因子的默认设置仅考虑血浆和血清，因此当运行其他样品基质时，这些稀释因子可能需要根据特定基质中预期的蛋白质浓度进行调整。样本制备指南可以应用于选定的部分基质。请联系support@olink.com了解更多信息。

当样品被送往Olink分析服务团队时，样品的稀释步骤将于分析前在内部进行。

对于组织或细胞裂解液，我们推荐的总蛋白浓度范围是0.5–1.0 µg/µL，浓度可以由BCA或Bradford测定。细胞也可以以cells/µL的浓度表示。为了确保每次分析所测量的蛋白质浓度都在最佳浓度范围内，建议在至少两种不同的起始浓度（即稀释）下运行样本。所有需要共同分析的组织或细胞裂解液具有相同的蛋白或细胞起始浓度水平是非常重要的。

检测范围

需要确定分析检测的范围以表征检测性能。为了了解Olink panel的检测能力和范围，每个panel的验证数据文档中都列出了每种测定方法的 定量下限（LLOQ）和定量上限（ULOQ）。这些信息可以确定该panel是否适用于特定范围。

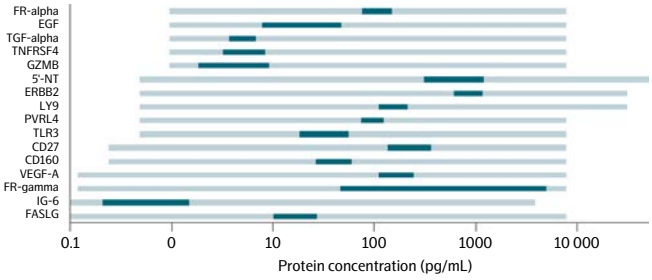

图2. 高亮部分显示出了血浆正常水平的动态范围。

数据共享与对比

数据共享

一个推动蛋白组学领域进一步提升统计学使用的方法是持续提高数据开放共享。 许多研究问题是由多个团队解决的，仅仅强调单个团队在统计上的显著发现可能会产生误导。

为了共享Olink数据，我们建立了几个合作框架，如SCALLOP（www.olink.com/scallop），它是一个发现和跟进基因与蛋白质关联（pQTLs）的联盟[7]。

想要成为SCALLOP联盟的成员，您需要成为Olink和全基因组基因分型数据收集的主要研究员。

在Folkersen等人的一项主要研究中，建议了用大样本量来检测基因和蛋白的关联[8]，这也说明了合作关系正变得越来越重要。

比较NPX值并与研究结合

为了比较研究间的NPX值，将研究标准化来消除他们之间的任何偏差是至关重要的。 如果研究可以被认为是随机的，或者研究之间的差异因素不重要，可以使用中位数中心化强度归一化。否则应采用桥接样本进行归一化，类似于前文中提到的纵向研究中结合不同时间点的数据。如有可能，这些桥接参考样本应具有研究代表性，如包括所有研究组，并应与其他样本的基质匹配。

更多信息参见白皮书*《数据的归一化与标准化》*[6]。

信息输出最大化

一般而言，通常会要求采用另一种技术对具有临床意义的主要发现进行独立验证，不考虑最初采用的方法。如果您需要对手中重要的Olink结果做此种验证，进行单重分析则是一个合适的选择。

已经有大量的案例表明了我们的92重 panel与单一的标准ELISA具有非常好的相关性，并展现了同等或更优的性能[9]。

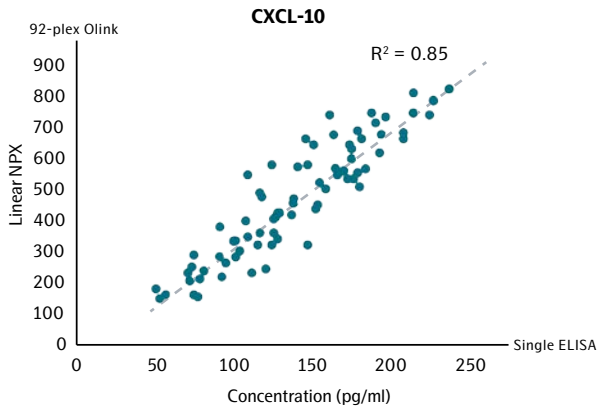

图3. 传统ELISA与Olink检测CXCL-10的相关性。

在对Olink检测与Quanterix的SIMOA检测进行比较时，也显示出了血浆样本中的良好相关性。

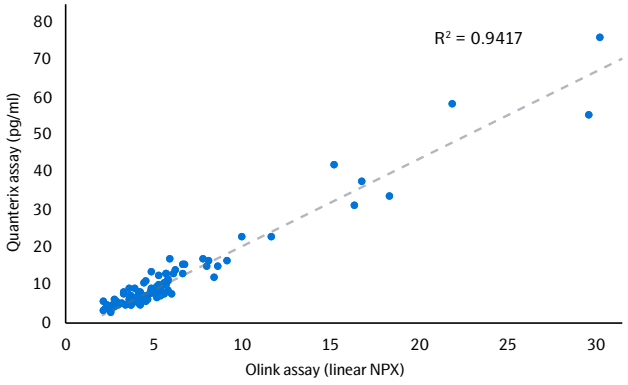

图4. Olink与SIMOA检测NFL的相关性。样品由Tomas Olsson (KI, Sweden)教授提供。神经丝轻链蛋白的检测采用NF-light™ 抗体（UmanDiagnostics，Umeå Sweden）。

Olink如何提供帮助

Olink的数据科学团队可在您的分析开始之前提供归一化方法和研究设计的讨论服务，帮助您从实验中获得最大收益。他们也可以帮助您定制化统计分析，将Olink panel的研究价值和信息输出最大化。

参考文献

1. Olink Proteomics, Pre-analytical variation in protein biomarker research, www.olink.com/white-papers
2. Dell *et al.*, Sample Size Determination. Ilar Journal (2012)
3. Bryan *et al.*, Multiplex screening of 422 candidate serum biomarkers in bladder cancer patients identifies syndecan-1 and macrophage colonystimulating factor 1 as prognostic indicators. Transl Cancer Res (2017)
4. Tromp *et al.*, Multimorbidity in patients with heart failure from 11 Asian regions: A prospective cohort study using the ASIAN-HF registry. PLoS Med (2018)
5. Yeh *et al.*, Assessing biological and technological variability in protein levels measured in pre-diagnostic plasma samples of women with breast cancer. Biomarker research (2017)
6. Olink Proteomics, Data normalization and standardization, www.olink.com/white-papers
7. Scallop consortium www.olink.com/scallop/
8. Folkersen *et al.*, Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. PLOS Genetics (2017)
9. Siegbahn *et al.*, A comparison of the proximity extension assay with established immunoassays in Advancing precision medicine: Current and future proteogenomic strategies for biomarker discovery and development. Science/AAAS (2017)

欧邻科生物科技（上海）有限公司

中国技术中心

地址：上海市浦东新区沪南路2157弄2号1011室
电话：021-50778771
邮箱：china@olink.com
网址：www.olink.com

版本号：1098,CN1.1, 2022-11-08

关注官方微信公众号
可获取更多详细信息