<http://iovs.arvojournals.org/article.aspx?articleid=2443689>

**更准确的青光眼视野进展建模：ANSWERS**

摘要

目的：验证视野（VF）进展分析方法，称为ANSWERS（非固定威布尔误差回归和空间增强分析），其考虑到随着青光眼进展而增加的测量变异性和测试位置之间的空间相关性。

方法：ANSWERS输出全局进展指标和每个VF位置的变化率的逐点估计。 将ANSWERS与平均偏差（MD）的线性回归和逐点线性回归（PoPLR）排列进行比较。 使用来自英国青光眼治疗研究所长达2年的视野系列。 这包括9104瑞典交互式阈值算法标准24-2 VF。 使用ANSWERS和PoPLR变化率来预测使用的距离基线7,13,18或22个月内子系列的下一次访视时的视野。 对预测未来VF的统计灵敏度，特异性和准确性进行了比较。

结果：在所有子系列中，ANSWERS检测VF变化的统计学敏感性明显优于MD和PoPLR的线性回归，特别是在短时间序列中。 ANSWERS的预测精度在所有系列长度上优于PoPLR，并且在较短的系列中改进特别明显。 相比于PoPLR，ANSWERS较好地预测了VF系列的75％。 ANSWERS的平均预测误差比PoPLR低15％。

结论：ANSWERS对于检测VF进展更为敏感，并且预测未来的VF损失优于MD和PoPLR的线性回归，特别是在短时间内。 （http://www.isrctn.com number，ISRCTN96423140。）

管理青光眼依赖于使用标准自动视野测量（SAP）测量的视野（VF），其评估主体是视野范围内的差异光敏度（DLS）。对于VF随时间变化的精确评估对于青光眼患者的适当临床管理是至关重要的，这样就可以使情况恶化的患者得到及时的治疗干预，而病情稳定的患者不会过度治疗。 然而，目前，VF测量是非常不精确的，具有复杂的统计特性，这使得监测VF的变化具有挑战性。

青光眼在VF系列中改变的临床评价可以通过分析算法得到支持。变化的概率和速率可从这些算法导出。这两个参数可以通过被称为趋势分析的方法来估计。逐点线性回归（PLR），是最广泛使用的趋势分析方法，为VF中的每个位置拟合一个普通线性回归模型，评估了拟合的意义和斜率。简易的测量，如测量与正常眼睛的平均DLS的偏差，也经常在趋势分析中使用; 但是，由于青光眼往往不会相同程度地影响所有位置，比起测量几个独特位置评估恶化的方法，全局指标通常具有统计灵敏度不足的问题。逐点的线性回归的置换分析（PoPLR），是逐点线性回归趋势分析的新发展，涉及一系列中的VF顺序的的随机置换。据报道，与PLR相比这种方法提供了改变的总体统计意义上更好的估计。PoPLR中改变的意义在VF的置换系列的研究范围内被估计，因此假定重新排序序列没有变化。因为在VF系列中需要置换，这种技术不能可靠地估计在少于六个VF的系列中变化的意义，因为在这样系列中有限数量排列可能性。此外，尽管由PoPLR估计整体统计学意义是不同的方法，底层回归模型仍然是普通线性回归，因此变化率和在个别位置变化的统计显著性的估计是与PLR一致的。

在VF测量中两个重要的内容不是解释当前用于探测VF系列变化的方法：非平稳变异性（增加的变异如DLS下降）和测试位置之间的空间相关性。视野测量受到重大变异的支配，这钟变异随着DLS恶化等级加重而增加，并最终会使盲区减小。例如，重复测量范围（90％置信区间[CI]）为7dB（26- 33分贝）时DLS是在32分贝健康，而此范围的增加至18分贝（5-27分贝）时DLS恶化到20分贝。随时间该改变的变异被称为非平稳测量变异。此外，视网膜神经纤维组成的VF的测试网格的遍历导致了在空间上相关的位置之间产生相关性。最广泛使用的SAP的VF测量，如那些汉弗莱视野分析器（HFA采取;卡尔蔡司，都柏林，CA美国）采用的，是在穿过病人视野的一个规则网格中进行。除了测试位置的邻域，该空间相关性也由视网膜神经纤维的解剖结构的制约。贝兹-Stablein等人对应于该视盘的六个扇区中的VF六个区域掺入这样的空间相关已经被证明了在检测VF进展的更高性能。因此，没有考虑到这些统计特性，用当前的方法进行VF改变的检测很有可能存在延迟或需要更多不必要的临床诊断。

一个新的趋势分析方法，分析非平稳威布尔错误回归和空间增强（ANSWERS），被提出并被从电子健康档案中获得的大量资料组验证。相对于常用的普通线性回归模型，即假定固定和正态分布的错误，ANSWERS包含了各级DLS的非平稳波动性，建模为威布尔分布的混合。这种使用贝叶斯框架的模型也包含了测量的空间相关性。尽管它的优化的统计属性，ANSWERS仍充当输出没有恶化的概率和在VF中特定位置的变化率，可以以和PLR同样的方式解释的线性回归模型。它也汇总本系列中所有的变化可能性产生一个总体的恶化指标，概述了一系列变化的总概率。

这项研究在来自一个临床试验的数据集上比较了ANSWERS与PoPLR和MD的线性回归。我们假定建模VF系列的更精确的方法应（1）在相同的特异性下检测变化更敏感，（2）更准确地预测未来的VF。

**方法**

这个方法是基于包括ANSWERS，PoPLR，和MD的线性回归的对比。此外，为了研究ANSWERS纳入空间相关性的效果，它的效果可以被关闭，该方法被称为ANSWER。

**数据集**

所有的VF都是采用24-2的测试模式和SITA（瑞典交互式阈值算法）标准测试算法的HFA（卡尔蔡司）测定。该试验在不包括两个盲点的52个测试位置测量视网膜DLS（动态光散射）。

在这项研究中，使用两个在不同中心收集的数据集。第一个数据集包含来自英国青光眼治疗研究(UKGTS)的VF系列,一项随机的、双盲安慰剂对照临床试验，测试与安慰剂相比，局部前列腺素类似物治疗,减少了VF恶化事件的频率的假设。患者被随访2年或者直到达到标准的终结点。在2年期间,患者在最开始后的第2、4、7、10、13、16、18、20、22和24个月都会测试一次,并且在最开始，以及第2、16、18、24个月都会重复两次VF测试。视野假阳性可靠性响应超过15%的测试已经被丢弃。只有得到了至少4个月(三次观察) 的系列才会包括在这个分析中。注意,系列的长度是纯粹是为了评价目的而不是ANSWERS所必需的。结果数据集包括437名患者的659个系列中的9104 次VF测试。随访时间的中位数(四分位距(IQR))是22(15 - 24)个月,这个系列中VF数量的中值(IQR)是11(6 - 12)。

第二个数据集是来自在英国伦敦眼科医院的管理下，对一群青光眼患者进行VF的重复测试可变性的检查。青光眼中视网膜功能的改变是缓慢的,假设在观测期间不会发生可测量的重大病变，就可以通过在短时间内重复测量来估计试验测量的可变性。27位患者的52只眼睛在短时间内(最多10周)被测试10次。这些重复测量中VF的方差表明测量固有的变化性。此外,每只眼睛的VF系列,和同一系列任意重新排序后的结果,表现出没有明显改变的稳定系列。在各种研究中，使用随机重排序的的系列来估计测量的可变性已经是一种成形的方法了。

病人的数据在调查之前都已经匿名处理了，不会包含个人或者敏感信息。数据保存在英国伦敦城市大学的一个安全的数据库中。因此,在这项研究中使用患者的数据并不需要他们的书面同意。这项研究遵循赫尔辛基宣言的原则,通过英国伦敦城市大学治理委员会的批准。匿名数据集可以按要求存取。

假阳性率和变化检测的统计灵敏度

比较了四种趋势分析方法的假阳性率和统计灵敏，假阳性是一个第一类错误，会在没有真正恶化的情况下检测到序列的改变。检阳性率可以在较短时间内重复测量获得的序列中估算出来。此外，把这些重复测量的数据随机重新排序产生的额外的伪序列中也没有产生真正的恶化。

在两次试验法的数据集中，来自每只眼睛十个视野的序列被随即重新排序300次。使用两次试验法中52只眼睛的数据集，产生了3到10个视野，共八种不同序列长度的124800(52 ×300×8)个伪序列。假设这些伪序列中每年采用其中的5个视野（这是UKGTS数据集的中值测量频率）。假阳性率会按照在进展中序列的比例来估计。临床情况下，假阳性可能会导致过度治疗和不必要的花费，因此，假阳性率高的方法在临床方面不会被认为是有用的。

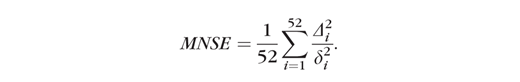
不同的方法应该在等价的假阳性率下做个依赖于选择的改变标准和序列长度的对比。PoPLR的恶化标准是整体的统计数据小于给定的一个临界值。ANSWERS和ANSWER的依据是恶化指标高于给定的临界值。MD的线性回归的依据是负斜率和斜率P值比一组临界值低。对于每一种方法，选定一组临界值来达到特定的假阳性率，然后在等价的假阳性率下比较每种方法的统计敏感性。

统计敏感性是识别真实改变的估量。理想情况下，敏感性应该被认为是在视野序列中检测到的进展和真实的潜在恶化以一定比例的组合。然而因为缺乏一个黄金比例和青光眼恶化的真实等级，每个视野序列的潜在恶化状态是未知的。因此，我们使用阳性率来进行各个方法的比较，阳性率就是UKGTS数据集中标记为恶化的序列的比例。给定一个未知的在数据集中真正恶化的序列比例P%，阳性率和统计敏感度有关，阳性率= (p %×灵敏度)+((1−p %)×假阳性率)。注意到如果所有的假阳性率都被控制为等价的，阳性率越高的方法敏感性越高。因此比较方法的阳性率就等价为比较方法的统计学敏感性。此外，当假阳性率比较低的时候，阳性率取决于敏感性。因此，在较低的假阳性率下比较两种方法的时候，阳性率的比例接近于敏感性的比例。这个比较是在基准期后的7,13,18和22月的测试序列做出的。

预测未来的视野

一个关于视野发展模型重要的临床问题是未来视野推测的缺失，这个是和视野改变速率紧密相关的。我们假设更好的进展模型会产生对变化速率更准确的估计，进而产生对未来视野更好的预测。通过在各个位置使用原始的DLS测试对ANSWERS，ANSWER和PoPLR进行了比较。注意，PoPLR的改变速率和预测就是对个体视野位置的普通线性回归。

UKGTS数据集的子系列被用来预测下一次访问时的视野，使用包括从基准期开始的前三次（四个月）访问的视野在内的最短的子系列。子系列的长度增加以便包含更多按年代升序排列的访问。对于每一个相同长度的子系列来说，使用三个方法来评估个别位置上的改变速率，然后用这个速率去预测下一次访问的视野。预报的性能评估为52个位置上评估与实际的视野的归一化均方误差(MNSE)：



MNSE是在UKGTS数据集中一致位置的预测误差平方与测量方差之比的平均值。在人口测量变化性的背景下量化预测错误的重要性，避免在较多易变的测量值的情况下高估错误，同时也避免在较少易变测量值的情况下低估错误。例如，在视野中心1db的错误比在周边视野中1db的错误意义更重大。因为在中央视觉的测量变化比在周边视觉要小。平均变化率（穿过视野位置的平均坡度）和预测性能的比较是在分别在所有子系列以及7、13、18和22个月的子系列中进行的。

三种方法在视野预测方面的绩效也在通过实际发生的变化数量研究，通过估计预测的视野和基准视野的平均差。所有的分析都是通过使用MATLAB R2013a进行的。

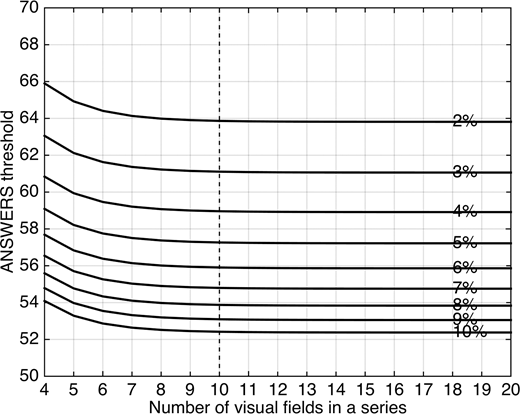
结果

UKGTS的视野数据和两次试验法的数据集的中值分别是-1.88（-3.83到-0.57）和-3.08（-8.36到-0.84）db.

变化检测的假阳性率和统计敏感性

任何一个特定的变化标准的假阳性率取决于视野序列的长度（在每个系列中，每一次附加试验都有可能产生假阳性率）。例如，ANSWERS对于不同的假阳性率和不同长度的序列（系列中的视野数目）的标准在图1绘制出来。临界值被拟合成一个关于假阳性率和系列长度的S型函数，拟合为r的平方>0.99。这就使得比在两次试验法中最大长度10还要长的系列临界值得以计算。

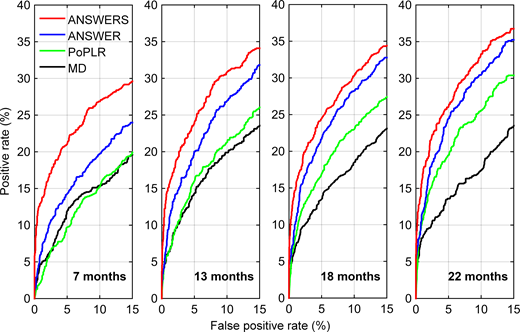
图1



ANSWERS以各种假阳性率和系列长度（系列中的字段数）更改标准。 ANSWERS阈值估计为假阳性率在2％和10％之间，系列长度在4和10之间。每条曲线表示在曲线结束时指示的假阳性率的ANSWERS阈值。 系列长度大于10的阈值（垂直虚线右侧的部分）是外推的。

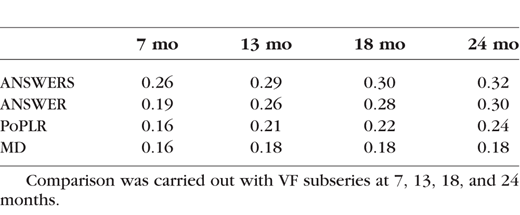
图2显示了从基线开始的7,13,18和22个月的系列的阳性率。 只显示0％和15％之间的假阳性率（特异性85％-100％）的阳性率。 表1总结了不同方法的部分阳性率曲线下的面积（图2）。 因为假阳性率在0％和15％之间的总面积为0.15，所以部分阳性率曲线下的面积通过除以0.15来归一化。

图2



从基线开始7,13,18和22个月时的VF子系列中ANSWERS，ANSWER，PoPLR和MD的线性回归的阳性率。 阳性率估计为0％和15％之间的假阳性率。

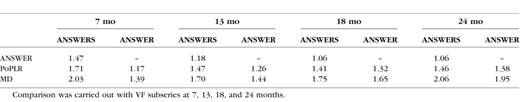
表1



ANSWERS，ANSWER，PoPLR和MD的线性回归方法部分正速率曲线下的归一化面积。

趋势分析方法也在5％假阳性率进行比较。 方法对之间的阳性率的比率显示在表2中，其中比率> 1表示更好的阳性率。 例如，在7个月的子系列中，ANSWERS与PoPLR的比率为1.71，表明ANSWERS的阳性率为PoPLR的1.71倍。

表2



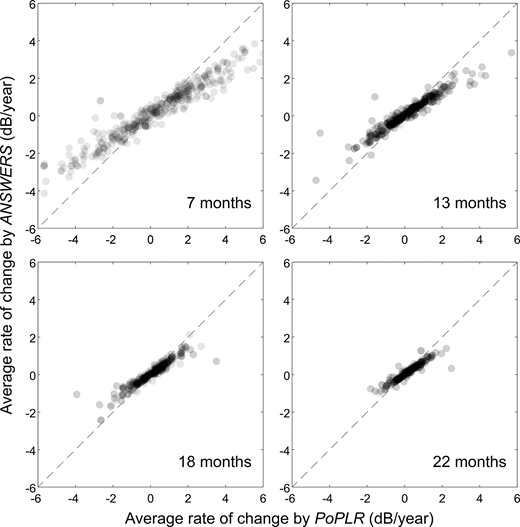
ANSWER和ANSWER的阳性率（假阳性率为5％）的比率超过ANSWER，PoPLR和MD线性回归的阳性率的比率。

在所有子系列中，ANSWERS和ANSWER的阳性率高于PoPLR和MD的线性回归。 短的子系列中的改善甚至更大。 与ANSWER相比，ANSWERS中包括的空间增强也增加了阳性率，特别是在短子系列中。 然而，随着子系列的长度增加，这种改善变得微不足道。

**变化率的估计**

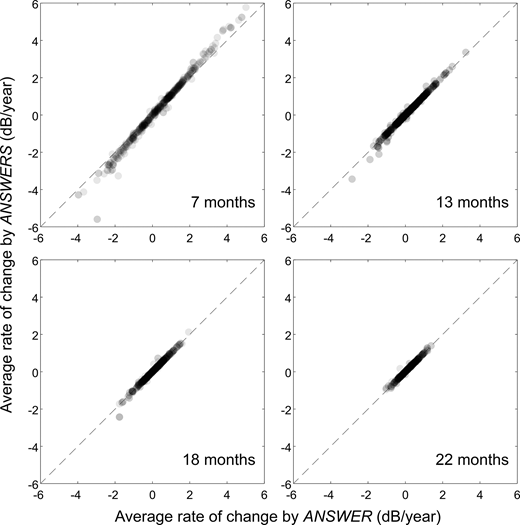
在所有子系列中，由ANSWERS，ANSWER和PoPLR估计的所有VF位置的平均变化率（中值[IQR]）为0.12（-0.44至0.67），0.12（-0.41至0.65）和0.16 (−0.73 to 0.98)dB /年。ANSWERS，ANSWER和PoPLR的平均变化率的（无符号）幅度（中值[IQR]）为0.55（0.26-1.07），0.53（0.25-0.99）和0.87（0.40-1.70）dB。ANSWERS和ANSWER与PoPLR相比都做出变化率变化幅度显著（P <0.01，Wilcoxon符号秩检验）较小的估计。与ANSWER相比，ANSWERS提供了统计学上显著（P <0.01％，Wilcoxon符号秩检验）更大的速率幅度。图3和图4中显示了子系列7,13,18和22个月中ANSWERS，ANSWER和PoPLR之间的变化率的比较，在所有子系列中，来自三种方法的变化率的幅度相关性 （PoPLR > ANSWERS > ANSWER, P < 0.01%，Wilcoxon符号秩检验）是一致的，除了第22个月，ANSWERS 和 ANSWER速率幅度（P = 20％，Wilcoxon符号秩检验）没有差异。 结果可以在图3中看出，其中点在具有小于1的斜率的线周围散射，在图4中，除了在22个月的那些点之外，点围绕具有大于1的斜率的线散射。 对于13个月和18个月，虽然ANSWERS和ANSWER之间的差异在统计上是不同的，但是差异的量是最小的，因此点紧密地散射到图4中的对角线。

图三



来自基线后7,13,18和22个月的VF子系列中ANSWERS的平均变化率与PoPLR的平均变化率的散点图对比。 半透明点用于减轻图中的重叠。 直虚线表示斜率为1的参考线。

图四

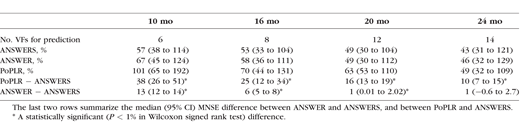


来自基线后7,13,18和22个月的VF子系列中ANSWERS的平均变化率与PoPLR的平均变化率的散点图对比。 半透明点用于减轻图中的重叠。 直虚线表示斜率为1的参考线。

**未来视野的预测**

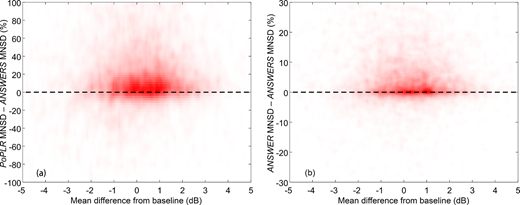
在VF的所有子系列中，ANSWERS，ANSWER和PoPLR的MNSE（中位数[IQR]）分别为54％（33％-113％），60％（36％-122％）和76％（43％ ％）。，ANSWERS分别在75％和71％的VFs中提供比PoPLR和ANSWER更好的预测（更低的MNSE）。 来自ANSWERS的MNSE明显小于（P = 0.01％，Wilcoxon符号秩检验）来自PoPLR和ANSWER的MNSE（中值[95％CI]差异分别是：15％[10％-19％]和2％[1％-4 ％]）。 三种方法使用7,13,18和22个月的子系列预测在10,16,20和24个月时的VF之间的比较总结在表3中。在所有子系列中，ANSWERS优于PoPLR。 在较短的子系列中的改善更大。 ANSWERS的空间增强使它在除了第24个月的所有VF预测中有比ANSWER更好的预测。

表3



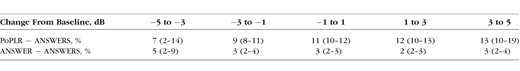
用于预测从基线开始10,16,20和24个月的VF的ANSWERS，ANSWER和PoPLR的中位数（四分位数范围）平均归一化平方误差（MNSE）

图5a显示出了与PoPLR相比，ANSWERS的VF预测的提高，其中PoPLR MNSE减去ANSWERS MNSD相对于变化量（预测的VF与基线VF之间的平均差）作图。 百分之七十五由ANSWERS预测的的VF（虚线以上）比PoPLR更好。 表4显示了当预测的VF与基准线有-5和5dB的不同之时，与PoPLR和ANSWER相比，ANSWERS的中值（95％CI）改善。 ANSWERS提供了更好的预测，而不管从基线的变化量。 因此，与PoPLR相比，尽管ANSWERS产生较小的斜率（较慢的变化率），但是通过使斜率趋于平滑，它在较快进展的眼睛中没有产生较大的预测误差。 此外，与ANSWER相比，ANSWERS中使用的空间相关性使其成为未来VF的更好预测因子，与基线的差异无关（图5b）。



ANSWERS比（a）PoPLR和（b）ANSWER的提高，通过相对于基线的变化量分层，确定为预测的VF与基线VF之间的平均差。 半透明点用于减轻重叠。 在y轴上的正值（虚线以上）表示ANSWERS有更好的性能。

表4



ANSWERS相对PoPLR和ANSWER从基线开始在VF中不同数量的改变的预测中的提高的中位数（95％置信区间）

**讨论**

ANSWERS是一种更灵敏的检测VF进展的方法，与其他基于趋势的方法相比，更准确地预测未来VF。在等效假阳性率下，与PoPLR和MD的线性回归相比，其检测到更大数量的具有变化的眼睛。此外，结果表明ANSWERS比其他方法可以更好地预测未来VFs。与普通回归模型中的正态分布误差相比，Weibull混合重测分布捕获了测量中变化的变量，并导致了显著的改进。此外，空间增强从VF中的空间相关位置收集信息，增加或减少在一个位置处观察到的进展权重，从而进一步改进该方法，特别是对于短时间序列。在经常由于有限的资源导致后续测试不常见的临床情况下， 在短序列中充分使用ANSWERS一定会有特殊的作用。

UKGTS数据的MD（中值为-1.88 dB）优于两次测试法的数据（中值为-3.08 dB），因此两次测试法的测试数据可能具有比UKGTS数据更高的测量变异性，因为随着DLS的减少增加的变异性。 因此，两次测试法的数据的假阳性率可能高于UKGTS数据的假阳性率。 这种潜在的假阳性率的过高估计会影响所有被等效评价的方法，因此它们之间的比较是公平的。

在这项研究中，两次测试法的数据用于估计测量变异性和假阳性率。这些数据是在非常短的时间内获得的（在不到10周内10个VF），并且在此期间不太可能发生可测量的损伤。然而，从其获得该数据集的患者可能比那些不像这么频繁的进行周围测试的临床患者更快地获得心理物理经验，因此从这些患者获得的测量值与在临床实践中获得的那些相比可能具有更低的变异性。然而，本研究中评价的方法使用相同的两次测试法的测试数据的共同点进行比较。如果两次测试法的数据集确实具有比试验数据集更低的变异性，特异性被高估，则阳性率可能被稍微高估。值得注意的是，未来VF的预测作为统计方法的独立验证，独立于两次测试法的数据。

所有在这项研究中比较的趋势分析方法假设在VF子系列的线性变化.这是因为没有足够的数据来确定非线性变化，如果存在的话，就是由于在临床实践中获得的VF系列相对较短。 最近的一项研究表明，VF系列的变化可能遵循非线性趋势，如指数函数. 然而，配置ANSWERS来模拟长VF系列中的非线性变化是很简单的。此外，PoPLR用于确定PLR进展的标准; 然而，对斜率和统计显著性的组合定义的其他标准是可能的。

重要的是要注意，由于测量的可变性，目前不能实现对未来VF的完美预测; 被预测的VF本身包括测量误差。 因此，统计方法的性能受到数据采集技术的限制。

总之，ANSWERS在建模VF进展中提供了一个“不确定性的风景”中的分析工具。 这种新技术有潜力帮助改善临床管理决策。 此外，它可以用于帮助定义更好和更相关的临床试验终点，这可以帮助提高试验的效率，减少其持续时间和成本。

**致谢**

由英国国家卫生研究所，国家卫生服务，支持的这项研究成员奖和健康技术评估基金支持这项独立研究。 本出版物中表达的观点是作者的观点，不一定是国家卫生局，国家卫生研究所或卫生部的观点。 资助者在研究设计，数据收集和分析，决定发布或准备手稿方面没有任何作用。

**参考文献**