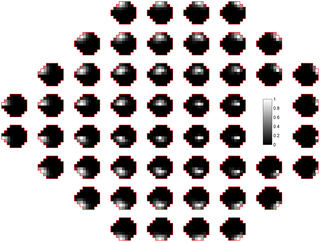
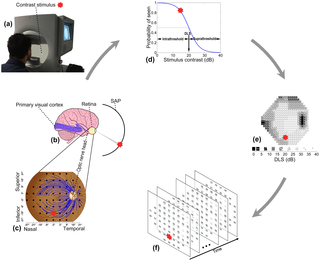
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0085654>

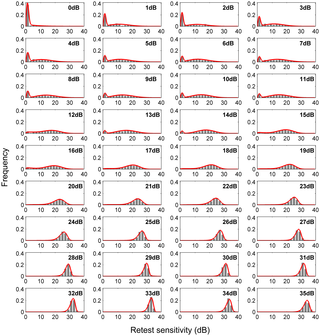
**检测视网膜功能的变化：使用非固定威布尔误差回归和空间增强的分析（ANSWERS）**

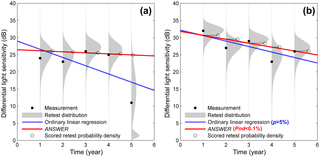
**摘要**

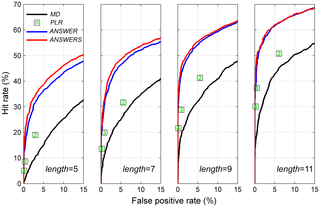
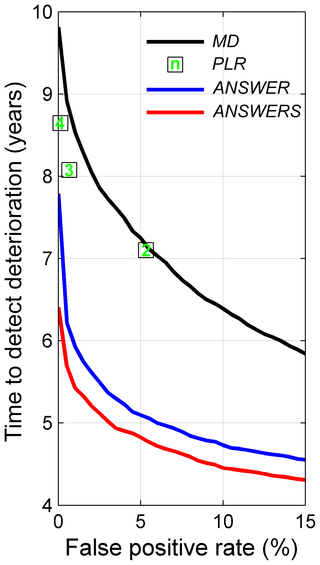
用标准自动视野测量的视野是用于确定眼部病理如青光眼中的视网膜功能的基准测试。它们随时间的监测对于检测疾病过程中的变化以及因此促进临床干预和在新疗法的临床试验中定义终点是至关重要的。然而，常规的变化检测方法没有考虑这些测量中存在的非固定测量变化性或空间相关性。提出了一种推理统计模型，表示为“具有非固定威布尔误差回归和空间增强的分析”（ANSWERS）。与通常使用的普通线性回归模型（假定正态分布误差）相反，ANSWERS将非稳态变异性模型化为与Weibull分布的混合。使用贝叶斯框架将测量的空间相关性也包括在模型中。它使用从电子健康记录获得的视野测量的大数据集进行评价，并与其他广泛使用的检测视网膜功能恶化的方法进行比较。 ANSWERS在匹配的假阳性率下能够比常规方法更早地检测到恶化。检测恶化的统计灵敏度也明显更好，特别是在短时间序列中。此外，与没有空间相关性的等效模型相比，ANSWERS中使用的空间相关性显示出提高了检测恶化的能力，特别是在短期随访序列中。 ANSWERS是检测视网膜功能变化的一种新的有效方法。它允许更好地检测变化，更有效的终点，并且可以缩短用于新疗法的临床试验的时间。

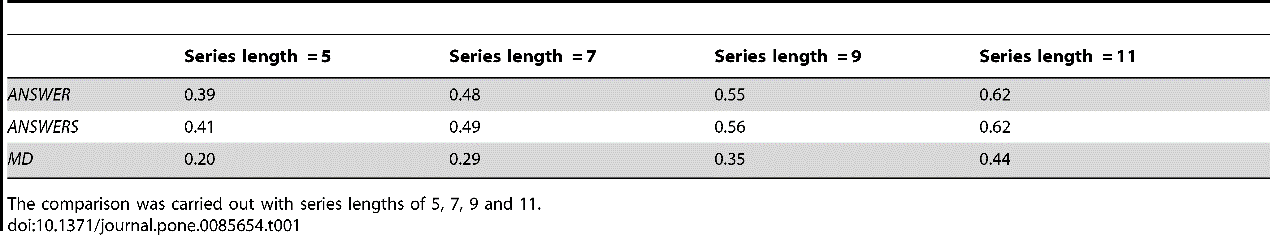
图表

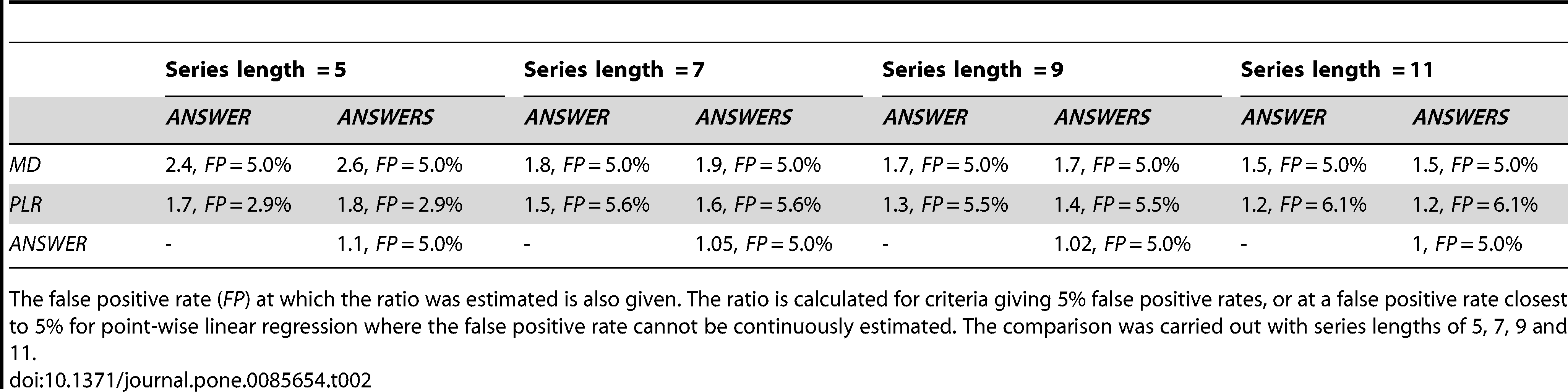












**背景和意义**

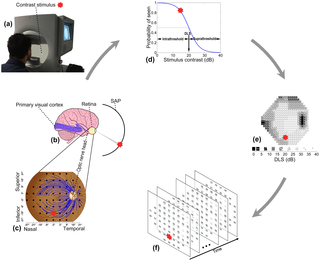
近年来，在研究实验室和体内了解眼部疾病方面取得了长足进步，导致神经再生过程甚至在某些情况下逆转失明的阐明，一个可访问和直接可见的大脑延伸，因此，视网膜研究正在成为解开其他神经系统变化的复杂性的焦点，如阿尔茨海默病，多发性硬化，和戈谢病。大多数眼睛病症的管理的主要目标是保持或改善视觉功能。对于视觉功能（即视野）建立的参考测试是标准自动视野测量（SAP;图1a）。 SAP通过人的视网膜和相应的视觉通路（图1b，c）测量差分光灵敏度（DLS）。不幸的是，用于分析来自SAP的数据的计算和统计方法的发展没有跟上眼睛相关研究的其他方面的进步。然而，SAP广泛地用于眼科和神经科诊所，特别是在青光眼的检测和管理中，青光眼是导致视网膜神经节细胞及其轴突进行性丧失并导致视网膜功能丧失的一组慢性视神经病变。这种疾病代表了一个巨大的全球健康问题，预计到2020年约8000万人将受到影响。通过用SAP测试监视视野来评估青光眼治疗的稳定性，在患者一生中以2个月至2年的间隔重复。 需要计算方法来分析SAP数据系列以识别变化; 已经发现如果没有这些，甚至有经验的临床医生都会做出不一致的决定。目前的统计方法通常在汇总的测量中，使用普通的随时间的最小二乘回归来追踪变量和感兴趣区域或个体视野位置。其他方法只是在最近的测试和基线测量之间进行比较。

图1.通过标准自动视野测量（SAP）测量的视野。

（a）来自SAP的对比刺激投射在视网膜的不同位置。当感知到刺激时，捕获来自受试者的反应。 （b）SAP评估视网膜和相应视觉通路的差分光敏性（DLS）。 （c）在视网膜上的各个位置（点）测量DLS。点（0°，0°）表示对应于视网膜上的中央凹的中心视力。视神经头是解剖盲点。测试位置不仅与它们的邻居相关，而且与通过它们的视神经纤维（其中一些被示出为蓝色曲线）相关。整个视野可以分为在垂直和鼻和水平上的颞区的上下视野。 （d）视网膜上的某个位置处的DLS以视觉系统响应于对比刺激的50％概率导出，并且与视觉途径中的中继神经元的光的生物反应相关。 （e）DLS以对数刻度测量，其在Humphrey Field Analyzer（Carl Zeiss Meditec Inc，Dublin，CA，USA）中计算为dB = http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e001其中，A是以阿普熙提为单位的的刺激的亮度，背景亮度是31.6阿普熙提。DLS范围在0dB（高对比度刺激，失明）和约35dB（低对比刺激，健康）之间，并显示为常规灰度图。 较深的阴影表示较低的DLS。 （f）DLS随时间的测量形成复杂的空间 - 时间时间序列。

<http://dx.doi.org/10.1371/journal.pone.0085654.g001>

当前用于检测DLS系列测量的变化的方法是不充分的，因为它们没有充分地处理数据的复杂性，特别是非稳态变异性和空间相关性。视网膜功能的SAP测量是间接的，因为涉及心理物理过程 - 人的反应取决于感知和响应光刺激的概率（图1d）。结果是相当大的变异性，随着疾病的进展，DLS恶化的增加，最终成为盲区审查。例如，当DLS在32dB处健康时，重复测量范围（90％置信区间）为7dB（26dB至33dB），而当DLS恶化到20dB时该范围增加至18dB（5dB至27dB）。这种随时间变化的变化被称为“非平稳测量变异性”。此外，SAP测量在患者视野内以规则网格进行，但是该网格不考虑从视网膜向脑传输信号的视网膜神经纤维的解剖结构（图1c）。视网膜神经纤维对网格的分割导致空间相关位置之间的相关性。有建模这种独特的空间过程的规定，但它们还没有被纳入SAP随时间测量的到系列分析中。因此，在不考虑这些统计特性的情况下，利用现有方法检测视网膜功能的变化通常会有延迟，或者需要比实际所需的更多的临床访问。

为了解决这些问题，我们提出一种分析方法来处理SAP数据中的变异性结构，并且还捕获关于支持视野变化的空间过程的信息。 这种新的计算方法，使用非稳态威布尔误差回归和空间增强（ANSWERS）进行分析，旨在准确识别随时间获取的SAP测量值的变化（图1f）。 此外，该方法可以适用于新疗法的研究，使得可以检测在干预之前和之后的变化。 在本研究中，我们将该技术应用于从电子健康记录中超过75,000名患者中采集的大规模临床数据。 具体来说，我们检查了这样的假设，ANSWERS可以比广泛使用的基于普通最小二乘线性回归的方法更快地检测视网膜功能的变化。

**材料和方法**

**伦理声明**

患者的数据在调查前被匿名化，并且不包含个人或敏感信息。 它在伦敦城市大学举行的安全数据库举行。 因此不需要这样的患者对他们的数据在研究中使用的书面同意。 该研究遵循了“赫尔辛基宣言”的原则，并得到了英国伦敦城市大学研究治理委员会的批准。 可以根据请求访问匿名数据集。

**数据集**

使用24-2测试模式（图1c）和SITA（瑞典交互式阈值算法）标准测试算法，使用Humphrey Field Analyzer（Carl Zeiss Meditec，CA，USA）通过SAP测量所有视野。 该测试在约50个测试位置测量视网膜DLS，其中每个测试位置在视野上均匀地分开6°的角距离（图1c）。

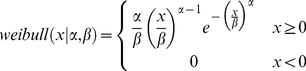
在本研究中使用在不同中心收集的两个数据集。第一个数据集是来自伦敦Moorfields眼科医院的青光眼诊所的电子健康记录的75,857名患者的402,357个视野中采集的。 DLS由于老化而恶化，并且通常不响应于用于青光眼的标准医学治疗而增加。因此，数据集中的所有系列都应该以至少等于年龄相关衰退的速率恶化。当观察到阳性率时，在青光眼的情况下，这通常是由于“学习效应”（患者学习进行视野测试）或测量的固有变异性。因此，每个系列的第一个视野被丢弃，以减少“学习效果”的影响。[26]，[27]如果在同一天拍摄多个视野，选择最后一次测量。在研究中只包括有在6年内获得并且包含至少7个视野的系列。注意，系列的长度纯粹是为了评估目的，并且不是所提出的模型所必需的。选择满足上述标准的所有系列用于本研究，并且所得数据集包括来自6,011只眼睛的6,011个系列的47,483个视野测试，代表约250万个单独的DLS测量。随访的中位数（四分位间距[IQR]）时间为9.3（7.9,10.4）年，每个时间序列中视野的中位数（IQR）为9（8,11）。视野测试之间的中值（IQR）间隔为1.0（0.6,1.4）年。

第二个数据集是来自一个研究，检查在Dalhousie大学，加拿大哈利法克斯在青光眼患者队列中进行的SAP的“测试 - 再测试”变异性。在青光眼中视网膜功能的变化比较缓慢。 通过在短时间内进行重复测量，可以在假设在观察期内不会发生可测量的恶化的情况下估计测量试验的可变性。[20] 在短期（最多8周）内对30名患者的一只眼睛进行12次测试，在此期间不会发生可测量的恶化。 这些重复测量中视野之间的方差表示固有的测量变异性。 此外，这些视野序列中的每一个以及具有任意重排序的相同序列表示没有潜在恶化的“稳定”序列。 使用随机重排序系列估计测量变异性是在各种研究中使用的已成型的方法。[28]，[29]

**计算模型**

用Weibull分布的混合来建模测量变异性。

单个DLS测量的可变性可以通过在短时间内重复视野测试来估计。[20]使用由1980（http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e003即30乘以12选择2组合）对的重复视野测试组成的测试重测数据集来估计范围从0dB到35dB的DLS测量的重测分布。重测分布通常是双峰的，截断的和偏斜的;由于DLS测量的非稳态变异性和删失性质，分布的形状在DLS测量的范围上显著变化。[22]由于重测分布不能由单个参数概率密度函数充分描述，在DLS的每个整数级，它被建模为Weibull分布的混合。选择Weibull分布是由于其多功能性和相对简单。与常用的高斯分布相比，它是一个更合适的选项，用于建模非负变量如DLS的概率分布。其概率密度函数由两个参数α和β定义：

 （1）

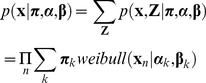
对于K个Weibull混合分量和N个重测数据点http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e005，潜在K维二进制矢量变量http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e006定义数据点http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e007属于哪个混合分量。 如果http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e009属于第k个分量，第k个元素http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e008，否则http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e010。 利用先验概率，

http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e011 （2）

被观察的和潜在的变量的完全可能性变为：

http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e012 （3）

其中http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e013，其中http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e014是http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e015属于第k个混合分量的先验概率，因此http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e016。http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e017，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e018，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e019，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e020其中http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e021和http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e022是定义第k个威布尔混合分量的参数。边界化（3）在Z上给出Weibull混合分布的可能性：

 （4）

（4）的最大化不给出参数，的闭合解，以及。 因此，导出期望最大化算法[30]以迭代优化（4）。 详细模型推导见附录S1。 此外，为了选择混合物组分的数目，K从1增加到在（4）中的似然性的对数在交叉验证中不再随统计显着性（p <1％）增加。

此外，由于在视野测试中的最小DLS的log Weibull分布，0dB，是未定义的，因此DLS被变换为：

http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e030 （5）

注意S(v)是v本身，除非当v<1，此时下限为0dB时。 该变换保证变换的DLS是连续的并且具有一阶导数，其是下一部分中描述的回归模型的优化的重要属性。

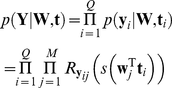
为了符号简单，DLS y的导出重测分布（4）在下文中表示为http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e036。

非稳态Weibull误差回归和空间增强分析（ANSWERS）。

我们提出了一种监测测量系列变化的方法，名为ANSWERS。 所提出的模型基于上文概述的Weibull重测分布的混合，并且在数据内并入空间相关。

在时间http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e037的时间序列中给出Q个视野测量（每个具有M个测试位置），Yij表示在时间ti和位置j的测量。 为了以紧凑符号表示回归模型，令http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e040，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e041，列向量http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e042和http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e043。

回归模型由测量中的M个测试位置中的每一个的权重向量http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e044定义。 每个权重向量Wj包含在线性模型的情况下随时间回归的第j个位置的斜率和截距。 为了简化符号，使用http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e046来指代所有权重向量的集合。 在时间序列Y中所有视野测量的可能性变为：

 (6)

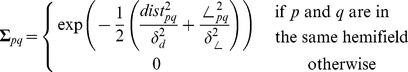
其中http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e049和http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e050分别在（4）和（5）中定义。 注意，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e051可以因式分解成i从1到Q的http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e052的乘积，因为http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e053有条件地独立于给定W的序列中的其他测量。

为了结合测量中不同位置之间的空间相关性，将斜率和截距的先验分布定义为多变量正态分布：

http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e055 (7)

其中Wa和Wb分别是回归线的斜率和截距。 Ma和Mb是各自正态分布的平均值，是由a和b缩放的协方差矩阵。

未缩放的协方差矩阵对测量中的测试位置之间的空间相关性进行编码。 第p行和第q列上的元素http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e062示视野中的点p和q之间的相关强度。 对于本研究中研究的视野DLS测量，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e063定义为：

 (8)

其中http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e065是视野中点p和q之间的欧几里得距离，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e066是视神经纤维穿过点p和q进入视神经头的角度之间的差异[23]，[24] http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e067和http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e068是选择为http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e069和http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e070的缩放参数。具体来说，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e071是视野中两个相邻点之间的距离，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e072是报告的进入视神经乳头的神经纤维进入角度的人口变异性的95％置信区间。注意，如果由于视神经纤维的生理分布，两个点位于视野的不同的半球（上或下;图1c）上，则http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e073。视野中的每个位置和所有其他点之间的非标准协方差在图2中示出。在本研究中，ANSWERS已经特别适于检测青光眼的恶化，因此空间相关性http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e075在此编码视神经纤维的解剖结构。为了使ANSWERS适用于对应于不同疾病或病症的其他类型的测量，应调整http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e076以反映该数据中存在的空间相关的特征。

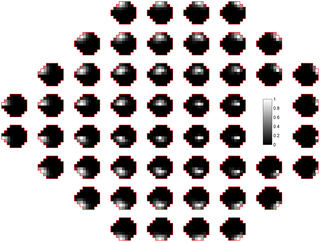


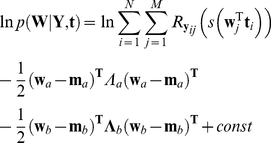
图2. 视野中每个位置和所有其他位置之间的空间相关http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e075。

图的组成是如图1c所示的24-2视野。在每个视野位置，具有24-2视野形状的图像表示该位置和视野中的所有位置之间的相关性。 在盲点位置处的灰度条表示相关性的水平。

http://dx.doi.org/10.1371/journal.pone.0085654.g002

选择尺度参数a和b的值以产生对于Wa和Wb的非信息先验。 确切地说，斜率先验被设置为Ma = 0（dB /年），http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e081对应于10dB /年的斜率标准偏差。设置截距先验，使得Mb = 18dB（DLS测量范围的中间）并且http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e083，对应于10dB的截距标准偏差。

根据（6）和（7），W的后验概率的对数可以推导为：

 （9）

其中，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e086，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e087。独立于W的项被归入常数项const。

后验概率（9）不能被识别为已知分布，因为（7）不是威布尔分布混合之前的共轭先验。 虽然对数后验（9）仍然可以相对于W最大化，但是难以在不知道底层分布的情况下估计W的精确方差。因此，如附录S1所述，使用拉普拉斯逼近[31]，[32]来近似http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e091作为以W模式为中心的正态分布。 拉普拉斯方法中的斜率和截距的估计精确匹配对数后验概率的局部最大值（9）。 然而，这些斜率和截距的方差是近似估计。

为了评估空间相关性的影响，可以通过将（7）中的http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e093的非对角线元素设置为0来“关闭”它的贡献。没有空间增强的模型被表示为ANSWER。

ANSWERS指数：变化的识别。

ANSWERS通过由拉普拉斯方法近似的方差估计斜率Wa和截距Wb。斜率的分布具有特别的临床重要性，因为它代表变化的速率和确定性。 “变化”同样适用于测量中的恶化（负变化）和改善（正变化）。 在进行性病症例如青光眼的情况下，每个位置的斜率分布可以被概括为“无恶化的概率”，其被量化为斜率≥0dB/年的累积分布。 “无劣化概率”值在下文中将被称为Pnd。 Pnd值的范围在0和1之间，其中较低的值表示较高的劣化概率。

为了总结视野系列中所有M个测试位置的恶化的可能性，全局指数，ANSWERS恶化指数http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e096定义为：

http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e097 （10）

其中http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e098是测量中第j个位置处的Pnd值。 http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e099是所有Pnd值的乘积的负对数，因此，非负值和较大值意味着关于测量序列的恶化的更大确定性。 类似地，为了评估测量系列的改进，例如在用于视网膜疾病的基因治疗的情况下，[3] ANSWERS改善指数http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e100可以推导为http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e101。 然而，因为本研究说明了鉴定青光眼视网膜功能恶化的应用，因此ANSWERS指数将仅涉及http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e102。

**评价**

为了评估ANSWERS在检测视网膜功能变化中的效用，有必要将其与目前在临床决策中使用的其它变化检测方法进行比较。逐点线性回归是最广泛使用的方法，其将普通线性回归模型与视野中每个位置的时间序列的测量相匹配，并评估拟合的意义和斜率。还使用诸如与健康眼睛的平均DLS的平均偏差的简单测量，但是由于青光眼倾向于不以相同的程度影响所有位置，因此与评估恶化的方法相比，全局指数通常具有不足的统计灵敏度而不能检测恶化[14] - [17]。此外，为了分别评估考虑非平稳变异性和空间相关性的益处，也评估没有空间增强的ANSWERS（ANSWER）。因此，ANSWERS与其他三种方法进行了比较：平均偏差的普通线性回归，逐点线性回归和ANSWER。

假阳性率的估计。

假阳性是I型错误，其中在没有真实恶化的系列中错误地检测到变化。可以在不足以发生可测量恶化的短时间段内获取的一系列重复测量中可靠地估计假阳性率。此外，随机重新排序这些重复的测量产生伪序列，其中也没有真正的恶化。

在测试 - 重测数据集中的30个眼睛的12个视野系列随机重排序300次，因此产生90,000个长度为3和12之间的伪系列。 假设在这些伪序列（在Moorfields数据集中的中值测试频率）中每年进行一次视野测量。 然后将假阳性率估计为鉴定为恶化的系列的比例。 在临床情况下，假阳性可能导致过度治疗和不必要的花费，因此具有高假阳性率的方法通常被认为临床上是没有用的。

应以相同的假阳性率比较不同的方法，这取决于所选择的变化标准和系列的长度。 对于平均偏差的普通线性回归，劣化标准是负斜率和低于设定阈值的p值。对于逐点线性回归，每个测试位置的劣化标准是负斜率和p值<1％，并且当至少n = 1,2,3和4个连续点恶化时视野恶化。对于ANSWER和ANSWER，恶化的标准是http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e103的值高于给定阈值。 对于每种方法，选择一组阈值以实现指定的假阳性率，然后在相等的假阳性率下比较每种方法的性能。

检测变化的时间。

使用Moorfields眼科医院的电子健康记录数据集，比较各个方法检测恶化的时间。 在每个视野系列中，包含前三个视野的子系列被认为是检测变化所需的最小系列长度。 然后通过按时间顺序递增地向子系列添加视野来增加子系列的长度。然后记录每种方法被标记为劣化的最短系列。如果通过方法在视野系列的任何子系列中没有检测到恶化，则将时间记录为系列的总时间跨度。不同方法之间的比较以相等的假阳性率进行。

命中率变化检测。

统计敏感性是方法识别真实变化的能力的度量。理想地，灵敏度应当被评估为具有真实潜在恶化的视野序列中检测到变化的比例。然而，由于缺乏对青光眼恶化的“金标准”和“绝对事实”分类[33]，每个视野系列的潜在恶化状态是未知的。因此，使用“命中率”来比较这些方法，就是在Moorfields数据集中标记为恶化的系列的比例。给定数据集中真实恶化序列的未知比例p％，命中率与统计灵敏度相关联：命中率=（p％×灵敏度）+ [（1-p％）×假阳性率]。注意，如果假阳性率被控制为对于所有方法是相等的，则更高的命中率意味着方法的更好的灵敏度。因此，将所有方法命中率的比较作为灵敏度的比较的替代。

结果

ANSWERS在MATLAB R2013a（MathWorks Inc.，Natick，MA）中实现。 对具有10个视野的系列的分析在2.50GHz的Intel i7处理器上花费大约1.5秒。 该软件可从作者免费获得。

**Weibull复分布的混合**

在所有级别的DLS中，将Weibull混合分量的数量增加到大于2不会显着增加交叉验证中的对数似然（4）。 因此，使用两种混合成分来建模0dB和35dB之间灵敏度的再测试分布。Weibull混合物在不同DLS下的直方图和导出的概率密度函数如图3所示。尽管有非稳定的变异性，但是每个分布可以通过两个Weibull分布的组合来充分描述。

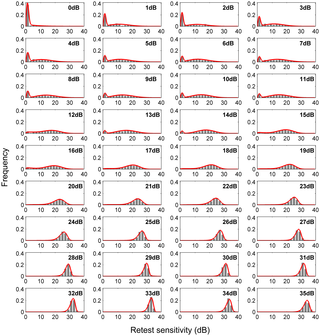


图3.

Weibull混合的推导概率密度函数以红色叠加。

http://dx.doi.org/10.1371/journal.pone.0085654.g003

图4中的示例表明了ANSWER使用的Weibull混合重测分布与单个视野位置处的DLS系列的普通线性回归相比较的作用。 因为为了说明的目的仅考虑单个视野位置，在这些示例中没有空间增强。 在图4a中，系列中的最后一次测量由于测量变化而突然改变，导致使用普通线性回归的陡坡。 通过比较，ANSWER较少受到最后一次测量的影响，因为它考虑了与在该DLS水平处的测量相关联的大变化性，因此产生较小的斜率。 ANSWER的这一性质使其对于DLS测量的非稳态可变性是鲁棒的，并且因此是变化率的更可靠估计者。

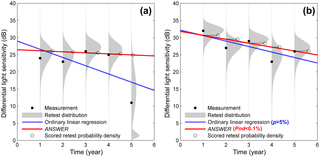


图4.比较ANSWER和普通线性回归的示例。

相应的差分光灵敏度测量的重测分布作为灰色区域叠加。 通过ANSWER回归线划分的概率密度在重测分布上标记。

<http://dx.doi.org/10.1371/journal.pone.0085654.g004>

**检测变化的时间**

尽管ANSWER有较强的鲁棒性，它并不降低检测恶化的灵敏度。 事实上，通过考虑DLS测量的非平稳变异性，该方法能够检测短时间序列中的显著恶化，在短时间序列中常规方法不能达到统计学显著性。在图4b中，普通线性回归没有表明显著恶化（p值> 5％），而ANSWER设法以高可靠性确定恶化发生（Pnd <0.1％）。 此属性可以使得ANSWER在检测恶化时提供更好的时间效率。

图5显示了每种方法在0和15％之间的假阳性率（具有较高假阳性率的方法在临床上是没有用的）下，首次检测视野系列中的恶化的平均时间。 因为逐点线性回归的标准（在视野中具有恶化的连续点的数量）不是连续的，不能使用连续假阳性率来估计逐点线性回归的时间效率。 此外，逐点线性回归的单点标准的假阳性率高于15％，因此在图中未示出。

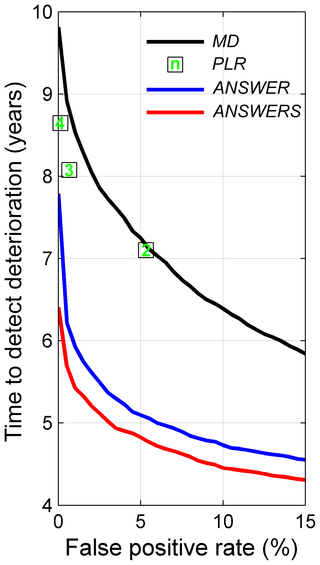


图5.在0到15％之间的假阳性率下，平均偏差（MD）线性回归，逐点线性回归（PLR），ANSWERS和ANSWER检测恶化的时间。

逐点线性回归中的连续点的数量在方形点中示出。

http://dx.doi.org/10.1371/journal.pone.0085654.g005

对于每种方法，将在5％的假阳性率或最接近5％的逐点线性回归率（两个连续点，假阳性率为5.3％）下比较检测变化的时间。 在这个假阳性率下，ANSWER比逐点线性回归（p <0.1％配对t检验）和平均偏差线性回归（p <0.1％配对t检验）更快检测到恶化。 此外，通过使用空间增强，ANSWERS能够比ANSWER（p <0.1％配对t检验）显著快速检测恶化。 平均来说，ANSWERS能在比逐点线性回归提前2.42（95％置信区间[2.35,2.49]）年检测到恶化，比平均偏差线性回归早2.28（95％置信区间[2.20,2.35]）年，比ANSWER早0.27 （95％置信区间[0.22，0.31]）年。

**变化侦测的命中率**

使用Moorfields数据集，使用各种长度的系列和在0和15％之间的假阳性率来估计四种方法的命中率。图6示出了系列长度在5,7,9和11时的命中率。仅显示指定的在0和15％之间的假阳性率下面的命中率（具有更高假阳性率的方法临床上不是有用的）。 在表1中比较了不同方法的部分命中率曲线下的面积（图6）。因为在0和15％之间的假阳性率的总面积为0.15，所以部分命中率曲线下的面积通过除以 0.15来标准化。 因为逐点线性回归的命中率不能用连续假阳性率来估计，所以不估计部分命中率曲线下的面积。

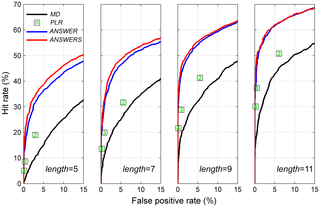


图6. 系列长度为5,7,9和11时，平均偏差的线性回归（MD），逐点线性回归（PLR），ANSWER和ANSWER的命中率。

逐点线性回归中的连续点的数量在方形点中示出。 命中率在0和15％之间的假阳性率下估计。

http://dx.doi.org/10.1371/journal.pone.0085654.g006

thumbnail

表1. ANSWER，ANSWERS和平均偏差（MD）的线性回归的部分命中率曲线下的标准化面积，。

<http://dx.doi.org/10.1371/journal.pone.0085654.t001>

该方法还以5％的假阳性率进行比较，或者在逐点线性回归（两个连续点标准）的最接近5％的比率下进行比较。表2中示出了方法对之间命中率的比率，其中比率> 1表示更好的命中率。例如，对于7个视野的系列，ANSWERS与平均偏差线性回归的比率为1.9，表明ANSWERS的命中率几乎是后一种方法的命中率的两倍。

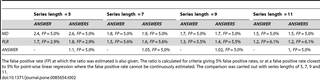


表2. ANSWER和ANSWERS（以列为单位）的命中率与平均偏差（MD）的线性回归，差分光敏感度的逐点线性回归（PLR）和ANSWER（以行为单位）的命中率的比率。

<http://dx.doi.org/10.1371/journal.pone.0085654.t002>

在所有系列长度中，ANSWER和ANSWERS的命中率高于平均偏差的线性回归和DLS的逐点线性回归。短系列有特别改善。这解释了ANSWER和ANSWERS为什么能更快速地检测恶化的效率。与ANSWER相比，ANSWERS中包含的空间增强也提高了命中率，特别是对于短序列。 然而，随着系列长度的增加，这种改善变得微不足道。

ANSWERS与其他方法相比的案例研究在附录S2中提供。

**讨论**

ANSWERS检测视网膜功能的变化比传统的统计方法更快，而不影响假阳性率。在等效的假阳性率下，当与通过其它广泛使用的方法检测的数量相比时，其还检测更多数量的具有视网膜功能变化的眼睛。 Weibull混合复测分布，与普通回归模型中假设的正​​态分布误差相比，使得ANSWERS获得关于退化状态的高确定性（图4b）。此外，空间增强聚集视野中相邻位置的信息以“确认”空间恶化模式，进一步改进该方法，特别是对于短时间序列。这种检测视野变化的空间元素以前很少被考虑过。[34] - [37] ANSWERS不仅可以帮助临床做出即时干预治疗的决策，而且还为眼科相关研究中的临床试验定义更有效的终点。 [3]

ANSWERS在短序列中的应用和有用性具有特别的临床意义。 由变化检测的普通线性回归代表的当前广泛使用的方法在短序列中受到限制，因为它们几乎不能达到所需的统计意义。 在临床情况下，往往由于资源有限，随访测试不频繁，这些标准分析可能会延迟检测视网膜功能的变化。反过来，就会延迟所需的治疗强化。在临床试验中，未能及时捕捉到改变也可能延长试验。

当选择ANSWERS检测视野系列恶化的阈值时，考虑所选http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e104的阈值的假阳性率是至关重要的。 在本研究中，阈值是从给定的假阳性率和每个视野系列长度的测试重测数据集估计。 然而，可以在理论上描述分析处方，并且在附录S1中提供。 注意，假定恒定假阳性率，http://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0085654.e104阈值不随系列长度改变。

在ANSWERS中使用的拉普拉斯方法在后斜率和截距分布（9）的模式下提供局部正态近似，因此这些回归参数的方差的估计可能不能捕获分布的每个特征（例如偏度）。 虽然真实的后验分布（9）是未知的，但是从拉普拉斯近似估计的斜率方差仍然被证明是相对于当前的其他方法在检测变化和量化变化的确定性方面的有效变量。

ANSWERS的开发理念是它可以适用于具有类似统计特性的其他应用，这在其他医学和生物测量中并不罕见。例如，用于预测肾衰竭的血清肌酐测量[38]，用于评估心脏病发作风险的心率测量[39]和糖尿病中压力感受器敏感性反馈 [40]在临床决策中提出了类似的挑战。有两个必要的步骤，以适应ANSWERS应用于其他类型的临床测量。首先，非稳态变异性应该从所讨论的测量中导出。由于Weibull混合分布是通用的和简明的，它可以使用附录S1中给出的期望最大化算法很容易地调整来建模其他重测分布。第二，当前空间相关性（8）源于视网膜神经纤维的解剖结构，因此不直接适用于除了视神经病变之外的测量和条件。因此，（8）中的空间相关性将需要被适应（或者如果必要，被去除）以反映所讨论的测量或疾病过程的空间特性。此外，ANSWERS用于推断线性变化，因为在临床实践中，由于短视野序列，通常没有足够的数据来识别非线性变化;然而，配置ANSWERS来测量显示非线性变化的长序列和时间过程的条件变化与是不重要的，可以通过改变（6）中的时间矢量ti到非线性组件，如径向基函数，来实现。

在这项研究中，测试重测数据用于估计变异性和假阳性率。由于缺乏关于视网膜功能恶化的黄金标准，这些数据是在非常短的时间内获得的（在不到8周内12个视野），因此在这段时间内不太可能发生可测量的损伤。然而，产生该数据集的患者可以比通常进行视野测试的频率更低的一般临床患者更快地获得心理物理体验。因此，测试重测数据中的患者可以产生具有比在临床实践中观察到的更低的变异性的测量。 然而，所有方法使用相同的重测试数据进行评估，因此每种技术的假阳性率将被等价地低估。 因此，尽管有可能低估变异性，但是测试重测数据确实允许我们在评估的方法之间做出公平的比较。

重要的是要注意，尽管新的用于分析视网膜功能变化的统计方法的不断演变，改进数据采集技术应该继续在研究的前沿。在测量采集点产生较少的可变数据将允许产生更精确的变化检测。 研究已经证明了各种改进DLS测量的方法。实例包括但不限于，调整刺激尺度，[41]，[42]以线性尺度而不是对数尺度[43]测试，以及增加测试点的密度或改变测试点的空间排列[44] [45]据报道，小于15 dB的DLS与神经节细胞的损失无关，并且可能不包含有关视网膜功能完整性的重要信息。[46] 因此，实际上需要在DLS的变化超过15dB的时候精确地测量DLS的变化。

总之，ANSWERS在检测视网膜功能恶化的不确定性的领域中提供了一个解决方案。这可以，例如，影响如何监测和治疗青光眼患者以及临床试验的效率和持续时间。在统计学敏感性和检测变化所需的时间方面，ANSWERS的表现优于常规的检测视网膜功能退化的方法。 ANSWERS被证明可以检测由青光眼引起的视野恶化，但是在其他具有非平稳变异性和空间相关性的测量学科中有很大的应用范围。

支持信息

附录S1。

详细的数学推导。最大期望算法威布尔混合分布，ANSWERS的拉普拉斯近似和计算给定假阳性率和系列长度的ANSWERS阈值分析模型。

doi：10.1371 / journal.pone.0085654.s001

（PDF）

附录S2。

示例说明ANSWERS与正在研究的其他方法的比较。

doi：10.1371 / journal.pone.0085654.s002

（PDF）