

# 验证治疗措施的公平

## ——高质量研究促进高质量卫生保健

Imogen Evans, Hazel Thornton, Iain Chalmers

### 序

本书内容有益于我们的健康,揭开事关生死的决策是如何制定的谜底,指出这些决策常存在严重缺陷,并引发了全球医生修正其临床实践方式的挑战。

本书不会引起不必要的恐慌;它非常钦佩现代医学所取得的诸多成果。其目的始终是期望改善,而非诋毁医疗实践。

我本人最初洞察到医学中根深蒂固的草率是在 20 世纪 80 年代。当时我应邀作为一名业外人士加入判断乳腺癌最佳疗法的共识委员会。之后发生的事让我感到非常震惊(当您阅读了本书的第 2 章后,您也许会同样震惊)。我们采用的证据来自一流的研究者和临床医生,但我们发现一些最有名的专家往往仅凭直觉或彻头彻尾的偏见决定治疗方案,而医生的决策和偏见将决定一名女患者生存的机会及被手术毁容的可能性。例如一位外科医生喜欢大刀阔斧的切除,另一位倾向于简单的肿块切除,而第三位则选择带有破坏性的放疗等。似乎这个科学评价的时代与他们相距甚远。

事实上,这样的情况过去常常存在,很多医生一直这样行医。尽管现在情况已有改善,但许多有才华、坦诚且技术娴熟的医生竟然不知道什么才是好的科学证据。他们各行其事,因为那是他们在医学院学到的,或因为其他医生也这么做,或因为这在他们既往经验中行之有效。但个人经验往往是可怕的误导——正如本书一针见血揭示的一样。

有的医生认为运用严谨的科学证据治疗个体病人是很幼稚的想法。他们声称,医学既是一门科学也是一门艺术。似乎很崇高,其实自相矛盾。医学知识有限,而任何个体的复杂性几乎无限,因此始终存在不确定因素,而在临床上,好的药物常需要正确推测。过去许多医学界人士常常混淆了推测与良好证据的区别。有时他们甚至轻言“确定”,即便事实上大可怀疑。他们回避应用可靠证据,因为他们不知道

应该如何评价证据。

本书阐述了在临床实践中个人经验与更复杂的临床决策之间的差异;如何以更好的方式区分有效或无效、安全或危险的治疗措施。尽可能采用通俗易懂的表达方式,比如“公平的试验”,警示科学与人类一切其他事件一样,容易出现误差和偏倚(人为或系统错误、自负虚荣、致命有害的药品、满足商业需求),但提醒我们即便如此,这仍是创造人类知识中几乎所有最突出进步的精细方法。医生(及媒体,象我一样)应停止诋毁临床研究,不要再视其为“人类豚鼠实验”;相反,应从道义上要求所有从业人员推动为病人而进行的公平试验,并要求病人参与其中。

对那些关注自身或家人健康及卫生政策的人而言,这是一本很重要的书。病人不仅是医疗保健的接受者,也是参与者。普通公众与医生及医学研究者肩负着同样的使命,因为我们既是医疗保健的接受者,也是医疗保健的支付者。如果我们只是医学被动的消费者,我们将永远无法提高医疗服务的标准。如果我们只喜欢最简单的答案,得到的将是伪科学。如果我们不促进治疗措施的严格验证,得到的将是无效甚至危险的治疗措施,将与真正有效的治疗措施擦肩而过。

本书以病人为中心,以改善医疗实践为目的,但对医生、医学生和研究者也同样重要,都可从中获益。最理想的是,能够强制每个新闻工作者阅读,并提供给每一位病人。因为如果医生不充分权衡科学证据,一切都会变得更糟,要知道我们的生死与这些科学证据休戚相关。

我保证:如果这本验证治疗措施的书对您是一个全新的话题,一旦您读完本书,您再也不会对医生的建议言听计从了。

广播电视节目主持人和记者

Nick Ross

2005 年 11 月 16 日

## 前 言

我们无法知道何时观察自然界的复杂事件才完整全面。Karl Popper 强调,知识有限,但无知是无限的。在医学上,我们永远不可能确定干预措施的结果,只能尽量减少不确定性。这种认识并非象它听起来那样悲观:具有挑战性的治疗措施,而后往往其的确证明正确可靠。这样的“工作真理”是合理坚固结构的基石,支撑着我们每天的床旁实践。

*William A. Silverman. Where's the evidence 1998<sup>[1]</sup>*

我们,本书作者已是多年的同事,不仅志同道合,而且都坚信医学治疗,无论新旧,都应当基于高质量证据。但卫生保健和研究的经验提示,事实并非如此,由此激励我们撰写了该书。

Imogen Evans 在其医学研究职业生涯中重视为病人所开处方的证据。当她成为《柳叶刀》的记者后,亲身体会到了一些药厂和研究者让人无法容忍的企图,即通过扭曲事实或美化自己的研究成果而谋取利益。Hazel Thornton 应邀参与一项质量可疑的临床试验,其意外经历使她意识到,她应该积极探寻自己的治疗进展,而非被动地接受。为确保医学研究有意义且能为病人提供高质量信息,她大力倡导医务人员与病人合作。Iain Chalmers 早在巴勒斯坦难民营工作时,就致力于严格评价医生给予病人治疗的疗效,他严格按照医学院所传授的方式行医,而他的一些病人却因此奄奄一息。此后,他一直大力提倡医疗保健决策应当基于当前无偏倚证据,患者应该知情决策;证据应来自相关临床研究,尤其是对照试验系统评价的结果。

每年有不计其数的研究探讨治疗效果,可悲的是,大部分研究都无视病人的需求,即便考虑了,其证据往往也不可

靠。我们希望本书可为读者进一步了解如何公正地验证治疗措施指明方向。本书不是针对个体治疗效果的最佳治疗指南,而是强调能够且应该进行确保基于问题的研究,以此回答与病人和医务人员相关的问题。

本书第一章描述了一些新的治疗措施怎样产生预期外的不良反应,而预期内的其他效应却未能实现。同时指出一些认为治疗措施无效的预测事实上是错误的,而一些有用的研究结果却未能用于实践。第二章强调医务人员尚未能充分评价许多常用治疗措施和筛检试验的公平性。第三章描述了一些“技术细节”,概述了公平验证治疗措施的基础,强调应减少潜在偏倚和考虑机遇因素的重要性,还介绍了诸如临床随机试验和安慰剂等概念,强调系统评价及相关证据的必要性。第四章描述了卫生保健各个领域的大量不确定性因素,及如何处理这些因素。第五章指出如何区分好、差及不必要研究治疗疗效的关键点。第六章介绍出于商业或学术目的如何扭曲已经完成的研究,而未解决可能真正影响病人疾苦的问题。第七章描述了病人可通过何种手段来确保获得更好的治疗措施。第八章勾勒了如何革新验证治疗措施公平性的蓝图及相关策略。

每章都参照精选材料撰写,其他信息来源列于本书末的附加资源部分。对希望获取更多细节问题的人来说,可以先看看 James Lind 图书馆,网址 [www.jameslindlibrary.org](http://www.jameslindlibrary.org)。该网址同时为读者提供了本书的联络信箱,欢迎读者将评论发至 [testingtreatments@jameslindlibrary.org](mailto:testingtreatments@jameslindlibrary.org)。

尽管我们描述了一些未充分验证的治疗所导致的危害,但绝对无意损害病人对医务工作者的信赖。我们的目标是促进互信与合作。但这需要病人帮助医生严格评价治疗的选择。我们希望读者能从本书中得到启示,不断提出一些关于治疗的尴尬问题,找出医学知识中的漏洞,共同参与研究以找出对我们每个人都有益的答案。

## 第 1 章 新的治疗措施不一定就更好,甚至可能更差

事实上几乎每周都有这样抢眼的新闻报道:某药物产生了意外副作用,某手术出现事故,某传染病猖狂肆虐,或某个妊娠治疗失误。一些批评家将当今基于科学的医学描述为没有人性的事业——似乎先于现代外科出现的屠宰式手术或曾作为治疗性药物的毒药在某种程度上都比现代医学更有人性。

然而不可否认,现代医学的确取得了令人瞩目的成就。有效药物的问世彻底改变了心脏病和高血压的治疗,还让众多精神分裂症患者走出精神病院,在家中接受治疗。有效治疗胃溃疡的药物让患者无需再经历大手术的折磨,同时让无效的治疗措施(如牛奶饮食)永远成为历史。对儿童进行免疫接种已使小儿麻痹症和白喉成为人们遥远的记忆。人们可能早已忘记,曾经认为患上白血病的人大都难逃死亡的厄运;而现在即便患上了其他癌症,也可以照常生活,并没有被病痛折磨致死。在西非和赤道非洲,有一种被称为“河盲”的疾病(即盘尾丝虫病),是由一种苍蝇幼虫叮咬所致,曾经令无数人失明,但通过药物治疗,该病现已基本根除。

现代成像技术也极大地造福于人类。超声、计算机断层摄影(CT)和核磁共振成像(MRI)技术有助于确保诊断的准确性和治疗的正确性。例如,MRI能够分辨卒中的类型,若卒中是由颅内出血所致(出血性卒中),那么可有效用于其他类型卒中的阿司匹林此时就可能变得十分危险。此外,外科和麻醉技术也得到了飞速发展。人造关节为无数患者带来了福音,器官移植如今也不再稀奇。

的确,由于社会和公共卫生事业的进步,如自来水、卫生设施、营养及居住条件的改善等,人们的健康状况得到了很大改善。就连对医学持怀疑态度的人也难以否认现代医疗保健事业给人类带来的不可磨灭的影响。在过去的半个多世纪里,人们的预期寿命延长、慢性(持续性)病患者的生活质量提高都在很大程度上归功于优越的医疗条件。

但即便是现在,许多医疗决策的制定并没有基于高质量的证据,现行的许多治疗措施仍对患者有害,而真正有价值的治疗却未能得到推广(参见下文和第5章)。设计新的治疗措施时,几乎总是存在对治疗作用和疗效的不确定性——极少出现具有压倒性说服力的明显治疗作用。因此,有必要开展设计精良的临床试验以确保治疗作用的评价切实可靠(参见第3章)。如果没有进行公平且无偏倚的评价,那么无用或甚至有害的治疗措施可能被认定为有效,而真正有用的治疗措施却可能被误认为无效而遭到摒弃。未经临床验证的疗效学说可能听起来令人信服,但实际上却远非如此。一些学说预言某治疗有效,但证据却提示无效;另一些学说断言某种治疗无效,但临床实践却证明其有效。

下面的事例已不是什么新鲜事了:18世纪,由于在长期航海途中坏血病夺去了大量船员的生命,詹姆斯·林德

(James Lind)医生开展了一项临床试验,比较治疗该病的6种方法。结果表明柑橘和柠檬(含维生素C)治疗该病非常有效。实际上,林德医生当时开展的就是现在的对照试验。

1747年,作为“HM Salisbury”的随船外科医生,林德选择了12名处于该病相似阶段的患者,将他们安置在船体的同一位置,并确保他们摄入了相同的基本膳食。这一点至关重要,因为这建立了“公平竞争的环境”(参见第3章、第4章)。每两个船员为一组,分别接受6种治疗之一以控制坏血病——苹果酒、硫酸、醋、海水、肉豆蔻,或两个柑橘和一个柠檬。最终,水果轻松地战胜了疾病,于是海军上将下令所有船只都要配备柠檬汁,结果到18世纪末,这种致命的疾病终于在英国皇家海军中消失。在林德医生所比较的治疗措施中,英国皇家医学院推荐的是硫酸,而海军上将则倾向于醋,但林德医生的临床试验却证明这两种权威观点都是错误的。而医学权威出错的情况并不少见(参见下文和第2、5、6章)。

如今,当医生和其他临床工作者就某种病情的最佳疗法持不同意见时,治疗效果的不确定性常会备受关注(参见第4章)。公众和医生对于消除这种不确定性起着重要的作用。为了符合患者和医务人员的最大利益,治疗措施的研究必须严谨。医生必须确保自己给出的治疗意见是基于高质量的证据,患者也应该要求确保他们所获取的治疗确实如此。只有建立了这一至关重要的合作关系,公众才会认同现代医学所提供的一切(参见第7章)。

### 查理二世之死

雷蒙德·克劳夫爵士(Raymond Crawford)(1865-1938)曾生动地描述了1685年英国国王查理二世的驾崩过程。查理二世罹患了脑卒中,御医们慌了手脚,仓促地开展了一系列野蛮的治疗措施:

“从国王右臂的一根静脉中放出16盎司血液,这立刻产生了不错的疗效。没有搬动国王而是让他继续躺在椅子上,这与现在的处理方法相符,但随后国王出现了惊厥。他的嘴被强行扳开,以防止牙齿咬到舌头。罗杰·诺斯(Roger North)将这一治疗方案简洁地描述为:先让国王清醒,再不让他入睡。国王的无数个御医闻讯后蜂拥而至,由于这次召见不分信仰和政治立场,所以他们都来了。御医们要求立刻在国王肩膀上拔罐,并施行深部划痕法,这样他们又成功放出了8盎司血液。他们还给国王喂下了高浓度的催吐剂(一种可引起呕吐的药物),可是国王只能喝下少部分。为了保险起见,他们又喂下了足量的硫酸锌和强泻剂,再佐以一系列的灌肠。他们还剪掉了国王所有的头发,将带有刺激性的起疱剂涂满了国王的整个头皮。这还不够,他们还要来一块烧红的铁块进行烧灼术。国王还在做死前忏悔,‘尽可能拖延临终时间’,但最后还是归天了。”

Crawford R. Last days of Charles II.

Oxford: Clarendon Press, 1909

## 意外的有害效应

过去,医生完全不清楚一种名为乙烯雌酚(DES)的人造(非天然)雌激素是否对有流产和死产史的孕妇有益。一些医生用了此药,另一些则没有用。20世纪50年代初,DES开始盛行,普遍认为其可以改善胎盘功能紊乱,这正是导致上述问题的原因。报道称有流产和死产史的孕妇经DES治疗后顺利产下活胎,这使得那些曾经用过DES的医生对此药的疗效更有信心。

例如,一位英国产科医生针对一名曾有过两次死产经历的女性患者开具了DES,并嘱其从妊娠早期开始一直服用。结果该妇女顺利产下一活胎。由此推断她成功分娩的“自身”能力经治疗后可能已经得到改善,因此在该妇女第4次妊娠时,这位产科医生没有再使用DES;但这次婴儿因“胎盘功能不全”在子宫内死亡。所以,在该妇女第5次和第6次妊娠时,这位产科医生和妇女都坚信应该再使用DES,结果这两次妊娠都成功产下了活胎。这位产科医生和妇女由此得出DES有效。不幸的是,此结论从未经过任何无偏倚试验的验证,却在此事发生后数年内广为报道。

更糟的是,大约20年后,一名年轻女子患上了一种罕见的阴道肿瘤,其母亲指出她女儿可能是因为是在妊娠期间服用了医生开具的DES才得了癌症。此后,大量研究显示无论男性或女性,若胎儿期曾在母亲的子宫内暴露于DES这种激素,均可能产生一系列的严重副作用,包括罕见肿瘤的发生率增加以及生殖系统异常。

到明文禁止DES用于妊娠期妇女时,已有数百万的男性和女性已使用了这种药物。试想:如果在20世纪50年代医生们早知道如何辨别当时可获得的最可靠的DES研究,就不会有这么多医生开具此药了。更严重的是,从未有人真正论证过DES对其主要适应症的疗效。人们普遍忽视了缺乏证据的严重性,让人不得不感到痛心。

### 婴儿流行性失明悲剧

第二次世界大战刚结束时,涌现出了许多改善早产儿预后的新疗法。几年之后,人们痛心地发现,许多婴儿治疗护理方面的改变显然产生了意料之外的有害效应。在这些临床悲剧中最突出的是,1942至1954年间出现的“流行性”失明,即晶体后纤维增生症。人们发现该病与用于治疗新生儿发育不全的补氧方式有关。12年来,为结束这场悲剧而付出的努力使人们清醒地认识到,所有的医疗变革在推广之前必须经过严格周密的评价。

*Sliverman WA. Human experimentation:  
a guided step into the unknown.*

*Oxford: Oxford University Press, 1985: Vii-Viii*

沙利度胺(反应停)是另一个治疗措施弊大于利的令人寒心的例子。20世纪50年代后期,这种安眠药作为比巴比

妥类药物(当时的常规处方药)更安全的替代物问世;与巴比妥类药物不同的是,过量的沙利度胺不会导致昏迷。鉴于沙利度胺还能减轻早孕反应,医生特别向孕妇推荐此药。

在20世纪60年代早期,产科医生发现一些新生儿出现了严重手足畸形的情况。这种医学史上罕见的疾病导致患儿肢体极其短小,好像手掌和脚掌是直接长出来的。德国和澳大利亚的医生认为这种婴儿畸形与其母亲在妊娠早期服用了沙利度胺有关。一名德国患儿的母亲对于这次发现起到了关键的作用——她报告称在其服用此药期间出现了手足麻刺和无力,这才引发了医生的重视。

1961年底,厂家召回沙利度胺。经过多年的公众游行和一系列的合法行为,受害者们终于开始得到赔偿。人们为这种毁灭性畸形所付出的代价是惊人的:在沙利度胺作为处方药销售(有的国家甚至将其作为非处方药销售)的大约46个国家中,成千上万名婴儿惨遭不幸。沙利度胺惨剧震惊了医生、药厂和患者,引发了世界各国对药物研发和审批过程的严格监察。

### 黄卡计划

在发生了沙利度胺致出生缺陷事件后,人们清楚地认识到上市后药物监测极其重要,因此,英国于1964年发起了黄卡计划。自那以后,约有40万份报告提交到了英国药品安全委员会(CSM),这个英国卫生部的下属机构,负责接收和分析这些报告结果。起初,只有医生才有资格递交报告,后来鼓励护士、药师、法医、牙医、放射技师和验光师都填写并递交报告卡。去年对黄卡计划进行了总结之后,上个月www.yellowcard.gov.uk网站开通了一个试行项目,现在患者和护工可借此在线报告疑似的不良反应。用户不仅可以在线填写报告表,还可以看到其他人递交的报告。尽管这些都是未经查证的疑似报告,但通过该计划能较好的把握各种药物的用药趋势。

*McCartney M. Doctor's notes. The Guardian.  
Health, 2005, Feb 8, p9.*

普拉洛尔(心得宁)事件虽不像沙利度胺事件那样家喻户晓,但也带来了巨大的危害。心得安属于 $\beta$ 受体阻滞剂类药物,用于治疗因心脏血供不足导致的心脏疾病,也可用于控制心律失常。第一代 $\beta$ 受体阻滞剂问世时,就有人警告说此类药物可加重哮喘患者的呼吸困难,故禁用于哮喘患者;还可能造成患者抑郁——“ $\beta$ 受体阻滞剂抑郁症”。经动物实验和简单的临床试验验证后,普拉洛尔成功上市,并作为一种较之前代药物作用于心脏的特异性更强的新药推广,由此认为其对于哮喘患者更安全。此外,普拉洛尔所导致的抑郁症也较少。总之,普拉洛尔似乎是极有前景的 $\beta$ 受体阻滞剂。

然而4年之后,在部分服用了此药的患者中,一种名为

普拉洛尔综合征的系列副作用越来越突出。主要表现为眼部并发症,如因泪液分泌减少导致的眼部干涩、结膜炎和因角膜损害导致的视力障碍。其他症状包括皮肤反应、耳聋和名为硬化性腹膜炎的严重疾病。发生硬化性腹膜炎时,正常的半透明腹膜变成了大量纤维化瘢痕组织,使肠和其他腹内器官出现绞窄。

事后看来,在普拉洛尔临床使用初期即有患者曾向其私人医生报告过眼部症状,但医生没有想到这类症状竟与该药相关。由于发现得太晚,造成的损失巨大——当1975年厂家召回由私人医生开具的普拉洛尔时,仅英国至少就发现了7 000名受害者。

30年来,针对药物试验的管理力度已大大增强,那么上述悲剧还会重演吗?毫无疑问,这种可能性已经减小,但即便是再完美的药物试验也不能绝对保证药物的安全性。普拉洛尔事件给我们上了生动的一课,至今仍启示人们:患者的观察结果和医生的警惕对于发现预期外的药物反应起着至关重要的作用。

### 预期效应未能实现

不要以为只有药物会害人,不当的建议同样可能致命。很多人大概都听说过美国儿童保健专家本杰明·斯波克医生(Benjamin Spock),他的畅销书《婴儿与儿童保健》(Baby and Child Care)被父母和专业人员都奉为宝典。但他给出的一条看似善意的建议却犯了极大的错误。这条建议从逻辑上看无可辩驳,并且显然有一定的权威性。他早在其1956年版的著作上就开始声称:“婴儿仰睡有两大坏处:第一,如果婴儿出现呕吐,他更容易被自己的呕吐物呛到;其次,婴儿倾向于将头偏向同一侧……这可能导致一侧头部扁平……因此,我认为一开始就让婴儿习惯俯卧位睡眠更好。”

让婴儿俯卧睡眠不仅成为了医院的标准规范,也令成千上万的父母在家中忠实执行。但现在我们发现,这种从未经过严格评价的建议导致了数以万计的婴儿猝死,而这些悲剧原本是可以避免的。虽然并非所有的婴儿猝死事件都归咎于这条错误的建议,但当否定了这条建议并提出相反意见即让婴儿仰卧睡眠后,婴儿猝死事件的确大大减少了。当20世纪80年代出现明确证据提示俯卧位睡眠有害时,医生和媒体立即开始警告俯卧位睡眠的危险性,之后婴儿猝死事件开始大幅减少。随后一项名为“躺着睡(back to sleep)”的齐心协力运动反复强调了俯卧睡眠的危险性,以尽可能消除斯波克医生那条令人遗憾的建议的负面影响。

尽管斯波克医生的建议看上去符合逻辑,但其理论依据却未经验证。类似的例子举不胜举。心脏病发作后部分患者出现了心律失常,相对于未出现心律失常的患者其死亡时间提前。因有药物能遏制心律失常,故推测这类药物也能降低心脏病发作后过早死亡的风险,这听上去合乎逻辑。但事实上这类药物的作用恰恰相反。临床试验仅评价了这类药物是否能遏制心律失常,但在1983年第一次系统评价不同试验积累的证据时发现,无证据提示抗心律失常药物能降低

### 基于未经验证的理论上建议可能致命

1956年斯波克医生所著的《宝贝的第一年》(A baby's first year)一书出版后,美国人广泛采纳了让婴儿俯卧睡眠的建议。大约十年后,欧洲和澳大利亚人也采纳了类似的建议。然而紧接着在上个世纪七、八十年代,SIDS(婴儿猝死综合征)的发生率急剧上升,其中俯卧睡眠的婴儿占了很大比例。如果1970年就对相关证据进行系统评价,当时就会发现俯卧睡眠的婴儿出现SIDS的风险比仰卧睡眠的婴儿高3倍。即便1988年首次公布了评价结果,但少有研究者意识到这些结果的重要性。直到上世纪90年代早期,人们发现在纠正了这一错误建议的地区SIDS的发生率下降了大约70%后,才在全国范围内发起了“躺着睡(back to sleep)”运动。在英国,这一错误建议在事隔21年后才得以纠正,至少有11 000名婴儿因此丧命。在美国,婴儿俯卧睡眠的现象更常见,盛行时间更长,死亡人数也更多。

摘自 Gilbert R, Salanti G, Harden M, See S. *Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of clinicians' recommendations from 1940-2000. International Journal of Epidemiology 2005; 34: 74-87*

死亡率。然而,这类药物在近10年的时间里仍被继续使用,继续导致患者死亡。在抗心律失常药物使用的高峰期,即20世纪80年代后期,有报告显示仅在美国,每年因使用抗心律失常药物导致提前死亡的人数就高达数万人,这在越南战争中死亡的美国人总数还要多。之后竟然发现,出于商业目的,某些提示抗心律失常药物可能致死的试验结果从未被发表过。

如果能减轻卒中患者的脑损伤,致残的几率肯定会降低。20世纪80年代,人们开始验证一种名为尼莫地平(属于钙拮抗剂类药物)的药物对于卒中患者是否有此疗效,一些动物实验得出了鼓舞人心的结果。当1988年发表的一项卒中患者临床试验显示尼莫地平对患者有益时,尼莫地平看似前途无量。然而,有关尼莫地平和其他钙拮抗剂药物的多项临床试验显示,试验结果互相矛盾。这可能是因为在卒中发作后早期给予这类药物,患者才能获益,针对尼莫地平试验的综述似乎也证实了这种可能性。但1999年,累计近8 000例患者的临床试验证据的系统评价显示,即使早期使用这类药物,也未发现其具有总体疗效。既然尼莫地平的使用似乎是基于充分的证据,为什么还会发生这种情况?鉴于人体研究的这一结果,第一次对动物实验结果进行了系统评价,结果发现动物实验的结果顶多只能算是不确定性结果。因此,在卒中患者中开展临床试验从根本上讲是没有充分依据的(参见第5章)。

对于更年期妇女常见的恼人的热潮红,激素替代疗法(HRT)疗效显著,且有证据表明HRT可能有助于预防骨质

疏松症。由于 HRT 所谓的有益效应越来越多,包括预防心脏病发作、中风以及其他效应,数百万的妇女在医生的建议下开始接受 HRT 长期治疗。但这样的主张事实上是站不住脚的。

单就心脏病发作而言,妇女们在过去 20 年间一直被告诫 HRT 能降低诸如心脏病发作这类严重疾病的风险,但事实上这一建议是基于有偏倚(不公正)的研究结果得出的(参见上文和第 3 章)。1997 年,芬兰和英国研究者系统评价了高质量研究的结果后警告,这条建议可能有误。他们发现,HRT 不仅不能减少心脏病,反而可能增加心脏病发作的风

### 难怪她会感到困惑

2004 年 1 月,一名子宫切除患者在给《柳叶刀》(The Lancet)杂志社的一封信中写到:

“我于 1986 年因子宫肌瘤而切除了子宫,医生同时切除了我的卵巢,并且发现我还患有子宫内膜异位症。因我当时仅 45 岁并可能因此立即绝经,我接受了激素替代疗法(HRT)。最初一年中,我采用口服结合雌激素(倍美力)治疗,但在 1988 年至 2001 年间,替我动手术的外科医生为我作了私人手术,每 6 个月对我进行一次雌激素皮下植入。我对这种疗法心存疑虑,因为一旦接受了雌激素植入,我感觉自己无法对其进行控制,而且术后数年我经常头痛。除此之外,我觉得自己很健康。

然而,我的私人外科医生向我保证,HRT 具有许多优势并且适合我,对此我也表示认同。HRT 已不像前些年那样仅被视为美容药物,有关其疗效的报道越来越多,不仅对心脏病和骨质疏松症有效,对于卒中也有部分预防作用。每次我去看病,我的私人外科医生似乎都能列出更多证明 HRT 有益的证据。

2001 年,我的私人外科医生退休后,我去找英国国民健康服务体系(NHS)的医生看病,他的话让我大吃一惊!他强烈反对我的私人医生关于 HRT 治疗有益的观点,他指出 HRT 将增加心脏病、卒中、乳腺癌的风险,也是导致头痛的原因,建议我停止 HRT 治疗。我继续接受了一次雌激素皮下植入,然后改为口服倍美力,也只服用了很短的时间,从那时起,我已有 8 个月未再接受 HRT 治疗。医生告诉我,由我来决定是否继续进行 HRT 治疗,为此我感到非常困惑……

我不明白为什么在如此短的时间内 HRT 及其显著优势就被全盘否定了。像我这样的外行要如何选择?尽管到目前为止 HRT 并未带给我太多的副作用,但我还是花了很多时间来讨论和思考我是否应该继续 HRT 治疗。整件事让我感到非常困惑,我肯定其他女性患者也会有相同的感受。”

Huntingford CA. Confusion over benefits of hormone replacement therapy.

Lancet, 2004; 363: 332

险。尽管一些著名的评论家否认上述结论,但两项大样本、无偏倚的临床试验结果证实了这一推测。如果 HRT 的疗效在其问世时就被正确评估,那么妇女们就不会被误导,也能避免许多人过早死亡。更糟的是,无偏倚的证据显示 HRT 可增加卒中和罹患乳腺癌的风险。

总的来说,HRT 仍是治疗妇女更年期症状的有效方法,但将其作为可降低心脏病发作和卒中的治疗方法来大力宣传,这一点十分可悲。虽然 HRT 导致出现这类严重疾病的风险仅稍有增加,但由于医生曾广泛使用该药,故受此影响的妇女实际上非常多。

未经充分评价的治疗措施即便不会导致死亡或危害患者,也会造成资源浪费。湿疹是一种令人烦恼的皮肤疾病,可累及儿童和成人。皮损处又痒又丑。尽管使用皮质激素软膏对此有效,但这种治疗的副作用令人担忧。20 世纪 80 年代早期,一种天然植物萃取油月见草油作为一种副作用较少、可有效替代皮质激素的药物问世。月见草油富含一种名为  $\gamma$  亚麻酸(GLA)的必需脂肪酸,故其用途貌似有理。例如有人认为,在湿疹患者体内 GLA 的转化(代谢)途经受损,故补充 GLA 在理论上对患者有效。琉璃苣油,又名七瓣莲油,因 GLA 的含量更高,也推荐用于治疗湿疹。

GLA 对人体无害但它是否有效?为此开展了大量研究,但得出的结果相互矛盾。厂商赞助的研究导致已发表的证据存在严重偏倚。1995 年,英国卫生部要求与月见草油厂商无利益瓜葛的研究者评价 20 项已发表和未发表的研究,结果无证据表明月见草油有效。因遭到厂商反对,卫生部从未将这份报告公之于众。但 5 年后,同一批研究者发表了另一篇关于月见草油和琉璃苣油的系统评价,结果显示,即使是大样本、高质量的研究也没有得出令人信服的证据证明以上治疗确有其效。

还有最后一种解释,即可能只有大剂量 GLA 才能起效。2003 年,一项严格实施的无偏倚临床试验将最后这种可能性推翻了。可笑的是,直到这些研究结果发表以后,英国药品管理局才最终在 2002 年 10 月,因无证据显示月见草油有效而撤消了这种昂贵药物的产品许可证。

同样重要的是,我们也不能被所谓针对致命性疾病的最新高科技成果冲昏头脑。某些细菌引起的重度感染可导致一种名叫感染性休克的严重并发症。这种并发症常见于本身患有基础疾病或免疫系统功能异常的患者。感染性休克患者的血压会降至非常危险的低水平,还会出现机体主要器官的衰竭。尽管加强了对感染的治疗,仍有 4/5 的患者可能死亡。

尽管细菌如何导致感染性休克的确切机理尚不清楚,但 20 世纪 80 年代发展起来的科学认识形成了一种推测,即认为该病与免疫系统功能失常有关。大多数严重细菌感染是由革兰氏阴性菌(指的是一种细菌分类的标准方法)所致,革兰氏阴性菌被视为导致感染性休克的主要原因,尽管某些革兰氏阳性菌有时也可能导致感染性休克。革兰氏阴性菌

### 我们这样做,是因为……

“我们(医生)这样做,是因为其他医生是这样做的,而我们不想和其他人不一样,所以我们也照样做;或是因为老师、同事和住院医生\*就是这样教我们的;或是因为老师、管理者、监督者、指南制定者要求我们这样做,并且认为我们必须这样做;或是因为患者希望我们这样做而我自己也认为应该这样做;或是因为更多的其他因素在诱导,我们认为应该这样做;或是因为对法律体系及对临床实践评估的担心,我们认为应该这样做;或是因为我们需要时间(让一切顺其自然),所以我们这样做;最后,我们也有失误的时候,因此必须做点事情为自己辩护。”

\* 住院医生(resident)的说法来自北美,住院医生(resident)在英国被称作“低年资医生(junior hospital doctor)”

*Parmar MS. We do things because.*

*British Medical Journal Rapid Response, 2004, March 1.*

可释放出一种名叫内毒素的毒性物质进入血液,刺激细胞释放出另一种名叫细胞因子的物质,而细胞因子可损伤毛细血管壁,导致这些在体内纵横交叉的小血管渗漏,引起休克和血压下降,从而造成感染性休克。

由此推论,如果能够清除血液中的内毒素和细胞因子,就能减轻感染性休克的严重程度。基于这一推论,科

学家们采用最新生物技术研发出可中和内毒素作用的特异性抗体。这些抗体首先在动物中进行了实验,并得出了鼓舞人心的结果:感染初期给予这些抗体可预防革兰氏阴性菌导致的感染性休克。然而,医生面临的困难在于当患者出现感染性休克时,无法立刻判断到底是革兰氏阴性菌还是革兰氏阳性菌感染,72小时后才能看到化验结果。尽管如此,报告称这个首个无偏倚(公平)临床试验的结果是成功的。

但不久,这一结果遭到了越来越多的怀疑。仔细审查试验结果后发现,以前对结果的诠释是不正确的。针对特异性抗体的后续临床试验均未能证实其有益效应,有时甚至提示其存在轻度负面效应。这些无偏倚研究得出的一致的阴性结果对免疫系统功能异常致感染性休克的科学理论提出了质疑,提示内毒素、细胞因子与感染性休克之间的关系远比原来设想的要复杂得多。人们最初对抗体疗法的热情大大减少,这也就不足为奇了。

### 要点

- 有偏倚(非公平)的研究可能导致不必要的伤害和死亡;
- 仅基于理论或专家意见本身,不足以证明治疗就安全有效;
- 系统评价无论对于临床试验和动物实验的设计和了解都必不可少;
- 患者对治疗中出现的非预期效果应引起重视。

## 第 2 章 已经使用的治疗措施 未经充分验证

在第一章中,我们了解到一些新的治疗措施产生了意外的有害效应,预期效应却未能实现;一些认为治疗措施无效的预言也被证明是错误的。本章重点讲述未经充分验证的治疗措施是如何的普遍。广告中常常宣传的乳腺癌治疗措施为我们提供了一些颇有价值的教训。

### 更多并不一定就更好

整个 20 世纪甚至直到 21 世纪,乳腺癌患者一直忍受着一些极其残忍和痛苦的治疗。无论手术或药物治疗,带给乳腺癌患者的伤害都远远超过其他真正需要承受的痛苦。然而这些却被患者和医生毫无疑问地接受了,患者深信越是激进的或毒性越大的治疗,越有可能战胜疾病。勇敢的医生和率直的患者代言人花费了数年时间才动摇了这种错误的观点。他们不仅要寻找可靠的证据来推翻“越多越好”的无稽之谈,同时还要遭受同行的嘲讽和医学权威的阻挠。

即便在今天,对疗效的担忧以及越多越好的观点仍然左右着治疗决策,促使一些患者及其医生选择“传统的”带有摧残性的痛苦疗法,而事实上并没有证据证明这些疗法就优于其他简单的治疗措施。为什么会这样呢?

#### 激进的治疗措施并非总是最佳的

“我们很容易认为,癌症治疗疗效好是因为采用了激进的治疗措施。为了保护患者免受不必要的风险和侵袭性治疗迟早要出现的副作用,采用随机对照试验对激进治疗和非激进治疗进行比较非常重要。这种比较是合乎伦理的,因为激进治疗既可能产生有益效应,也可能带来不必要的伤害,没有人知道最终会是什么结果。”

Rees G, ed. *The friendly professional: selected of Thurstan Brewin*  
Bognor Regis: Eurocommunica, 1996

直到 20 世纪中叶,手术仍然是乳腺癌的主要治疗方法,这是因为人们坚信乳腺癌进展缓慢且有序,如首先从乳腺肿瘤部位侵袭至周边腋窝淋巴结。由此推论,手术范围越大,手术时间越早,越有可能阻断肿瘤扩散。治疗措施基本上就是广泛的“局部”手术,即乳腺或乳腺周围组织的切除。虽然称为局部手术,但根治性乳房切除术要切除大面积的胸肌并广泛清扫大量的腋窝淋巴结组织。

然而,一些善于思考和观察的乳腺癌专家注意到这些日益剧增的摧毁性手术似乎并未对乳腺癌的病死率产生任何影响。于是,他们提出了一种新的理论:乳腺癌实际上一开始就是一种全身性疾病而不是通过周围淋巴结有规律地从乳腺开始扩散,换言之,他们认为在发现乳房包块时癌细胞就已经存在于身体其他部位了。如果真是这样,他们建议切除肿瘤及周围适当范围内的正常组织,再加上一个疗程的放

疗,会更有利于患者,并且可能与根治术同样有效。此时“全身疗法”的引入,即治疗对抗身体其他部位癌细胞的生长或繁殖,也是基于这个乳腺癌扩散的新理论。

这一新思路导致的直接后果是,医生开始主张采用范围更小的手术治疗,如乳房肿瘤切除术,即切除肿块及周围正常组织边缘,然后行放疗,部分患者采用化疗。但在比较新方法与根治术时,他们遇到了巨大的阻力。一些医生坚信根治术或其他疗法,患者也如此强烈地要求。其结果是大大延误了收集关于新旧疗法优缺点的重要证据。

尽管面对这些困难,过度手术最终还是受到了来自不愿再继续施行疗效不确定的根治术的外科医生以及不愿再忍受摧残性手术的直言不讳的患者的挑战。

20 世纪 50 年代中期,美国外科医生 George Crile 开始公开指出他对“越多越好”观点的质疑。Crile 医生认为没有其他办法可以激起医生的批判性思考,因此他在《生活》(Life)杂志中撰文呼吁道:医学界内的争论现在应是公开的,是可以公之于众的。另一名美国外科医生 Bernard Fisher 与其他专业的同事共同设计了一系列严谨的研究癌症生物学的试验,结果表明即使是在发现原发癌之前,癌细胞可能已经通过血液循环广泛转移。因此,如果癌细胞已经存在于身体其他部位,扩大侵袭性手术的范围也无济于事。

#### 传统的根治性乳房切除术

在 20 世纪中后期之前,William Halsted 爵士于 19 世纪后期提出的根治性乳房切除术一直在乳腺癌手术中颇为盛行。施术者不但要切除整个乳房,还要切除覆盖胸壁的胸大肌。为了便于施术者进入腋窝清除淋巴结和周围脂肪组织,甚至连胸小肌也会被切除。

#### 范围更大的根治性乳房切除术

在当时,“越多越好”的观点促使根治性手术的施术者进一步扩大手术范围,锁骨下淋巴结和胸骨下乳内淋巴结都被切除。为了触及乳内淋巴结,还会取掉几根肋骨,并用凿刀切开胸骨。一些外科医生觉得这样还不够,甚至还要切除患侧手臂和体内各个部位的腺体(肾上腺、垂体、卵巢),以此来抑制激素的生成,因为激素被视为肿瘤扩散的“动力”。

如果某位患者历经这样的手术而幸存下来,她剩下的将是一副任何外衣都难以遮掩的受到了极度摧残的胸廓。如果手术部位在左侧,那么只会剩下一层薄薄的皮肤盖在心脏上。

摘自 Lerner BH, *The breast cancer wars: hope, fear and the pursuit of a cure in twentieth-century America*. New York: Oxford University Press, 2003

Crile 医生是根据其临床判断倡导采用小范围的局部根治术,而 Fisher 医生与一个日渐壮大的研究小组则以更规

范、更严谨的方式展开了合作。他们试图通过众所周知的无偏倚(公平)的方法——随机对照试验来肯定或否定根治性手术的价值。他们认为通过这样的研究可以让医学界和公众信服。1971年,直言不讳的Fisher医生指出,所有外科医生都应出于伦理和道德责任通过开展这样的试验去检验其各自持有的理论。在Fisher医生的试验长达20年的随访结果显示,过早死亡的风险这一结局指标上,乳房肿瘤切除术加放疗与乳房全切术的疗效相当。

同时,在英国,Guy医院的Hedley Atkins及其同事在20世纪60年代早期开展了第一个比较乳腺癌保守治疗与传统的根治性乳房切除术的随机对照试验。结果同样显示,确诊后20年内这两种治疗导致的结局几乎没有差异。与英、美两国一样,瑞典和意大利也开展了其他随机试验,对其他治疗方法进行了比较,如单纯手术与术后加放疗比较,长期与短期化疗比较。

截至1985年,乳腺癌试验数量的不断增加使医生很难及时跟进所有试验结果。为了解决这一问题,牛津大学的Richard Peto及其同事将所有试验结果汇总,完成了第一个包括各试验所有受试者所有信息的系统评价。这样肿瘤专家和公众就能获取最新的合并后的全球证据。现在,乳腺癌治疗的系统评价正在定期更新并发表。

然而,摧残性手术的消失并未导致“越多越好”心态的终结,事实上远非如此。20世纪的最后20年内,大剂量化疗加骨髓移植或干细胞移植的治疗方法非常盛行。1999年,《纽约时报》(New York Times)的一篇评论文章总结了这种疗法背后的原理:

“医生从患者身上抽取一些骨髓或红细胞,然后给予足以破坏骨髓的大剂量毒性药物。希望以此清除癌细胞,待留取的骨髓重新输回体内后可以尽快生长,以免患者死于感染。通过骨髓捐献的方式对血癌患者进行骨髓移植,其疗效早已确立,但这是因为癌细胞存在于骨髓中。而要采用这种方法来治疗乳腺癌却是基于一种全新的、从未经过检验的原理。”

尤其在美国,成千上万的处于绝望中的妇女要求医生和医院为其施行这种非常痛苦的治疗,而医生和医院对此乐此不疲。每100位患者中会有5位死于这种治疗。该治疗通常要花费掉上万美元,并且部分需要自费。最终,保险公司迫于压力只好赔付了部分患者的治疗费,尽管并无证据支持这种昂贵的治疗。许多医院和诊所却从中牟取了利益。1998年,一家医院的年收入为1.28亿美元,其中大部分都来自于施行骨髓移植的癌症中心。对于美国医生而言,这不仅是一份名利双收的肥差,还是一个生产论文的富矿。执迷不悟的患者更是活跃了这个市场。美国的私立医院争相开展这项治疗,竞争相当激烈,甚至打出了降价广告。20世纪90年代,在美国,即便是募集患者开展临床试验的学术性医学机构也开展了这项治疗。这些值得质疑的疗法竟然成了癌症治疗机构的“摇钱树”。

无限制地开展这种未经证实的治疗还导致了另一个严

重后果:找不到足够的患者参加将该疗法与标准疗法进行比较的临床试验。因此耗费了比预期多得多的时间才得到了可靠的答案。

尽管在这种压力之下要获得无偏倚的证据困难重重,部分临床试验仍得以实施,其他证据也得以严格评价。2004年,对作为乳腺癌常规治疗的大剂量化疗加骨髓移植的累积结果进行了系统评价,结果显示尚无可靠证据证明该疗法有用。

#### 竭力获取无偏倚证据

研究者预计用3年时间募集约1000名患者参与两项试验,而事实上却花费了7年……这其实并不令人吃惊……患者进入临床试验时必须签署一份知情同意书,说明可能预后不良以及尚无证据证明骨髓移植优于其他标准治疗。进入试验之前,患者必须面对这些现实,这对于任何人都不是一件轻松的事。但如果患者没有参与设有对照组的随机试验而接受移植,热心的医生会告诉她骨髓移植可能挽救生命。虽然患者有权知道真相,但可以理解没有人愿意去找那些不给予她们任何希望的医生。

摘自 Kolata G, Eichenwald K. Health business thrives on unproven treatment, Leaving science behind. New York Times Special Report, 1992, 2 October

#### 筛查正常人群, 早期发现疾病

为早期发现疾病而筛查正常人群,这听起来是一种很明智的做法,但怎样才能更好地避免疾病的严重后果并保持健康呢?虽然有些疾病,尤其是癌症,已列入国家筛查项目,私人诊所也纷纷开展定期体检,其实就是一系列的筛查试验,声称这有助于就诊者保持健康。然而一些筛查项目是有益的,如测量血压,而一些却是有害的。

因此,在茫无目的地接受泛滥的筛查之前,应该先想一想筛查的目的是什么。个人或人群的筛查主要是通过检查找出可以从治疗中获益的人,以此来减少死亡和严重残疾的风险。早在1968年,WHO的一份报告就总结了评价筛查试验价值的基本标准,这些标准至今仍值得铭记:

- 要筛查的疾病应该关乎重大的健康问题
- 应有对疾病有效且可以接受的治疗措施
- 发现异常后应有相应的设备进行诊治
- 疾病的早期应该可以识别
- 筛查试验应该有效
- 筛查试验应该为人群所接受
- 应对疾病自然史有充分的了解
- 筛查造成生理或心理伤害的几率应小于其产生效益的几率
- 筛查应是一个持续的过程,而非一劳永逸
- 筛查项目应该价有所值

### 从普通人到患者

筛查不可避免地会将部分检测结果呈“阳性”的人转变成患者,但这样的结论不能轻下。“如果一个患者向医生求救,医生应竭尽所能。医生不对医学知识的缺陷负责,但如果医生开展了筛查项目则另当别论。照我们看来,医生应该有结论性的证据证明筛查可以改变大部分被筛疾病的自然病程。”

Cochrane AL, Holland WW.

Validation of screen procedures.

British Medical Bulletin, 1971; 27: 3-8.

如今,事后来,上述原则主要存在三大问题:第一,未充分强调调查的有害作用。几乎所有检查都会有风险,从某种意义上说,筛查试验检测疾病的能力是有限的。例如,筛查可能无法检测出所有或大部分患病患者,即敏感性不高;也可能引起过度诊断,即特异性不高。人们一旦被贴上了疾病的标签,常常就要再接受进一步的检查,随之而来的是焦虑,无端的歧视,如来自保险公司。第二,上述标准强调应对疾病有效且可以接受的治疗措施,但当前许多已被接受的治疗措施其价值并未得以证实。因此,基于这些并不完善的筛查试验结果给出的推荐治疗不可避免地存在不同程度的风险。第三,上述标准没有强调引入筛查项目的决策应基于高质量证据。

### 如何评价筛查试验的利与弊?

“人们对于筛查试验利弊的评价有所不同。例如,孕妇对于唐氏综合症的筛查可能会作出不同的选择:如何权衡生下一个先天愚型的小孩与羊水穿刺术致医源性流产风险之间的利害关系。

选择接受筛查的人认为自己从中受益了,而其他没有接受筛查的人也认为自己从中受益了。除非他们能够获取有关筛查试验利弊的高质量信息,并且能够正确地评估这些信息,才可能作出正确的选择。”

Baratt A, Irwig L, Glasziou P, et al.

Users's guide to the medical literature. XVII

How to use guidelines and

recommendations about screening.

Journal of the American Medical Association,

1999; 281: 2029-33.

从目前的筛查项目中我们可以得到什么教训呢?神经母细胞瘤(一种主要累及青少年的罕见恶性肿瘤)的筛查经验就很有启示意义。该筛查项目听起来很诱人,理由如下:(1)一岁前诊断,患儿的预后将优于一岁后诊断;(2)疾病晚期的患儿病情进展比疾病早期的患儿快得多;(3)该筛查试验既简单又便宜,只需要从湿尿布中采集尿样进行检测即可;(4)该筛查试验可检测出90%的神经母细胞瘤患者。

### 循证医学遭到质疑

1999年,一位在美国接受了家庭(全科)医疗培训的医生接诊了一位前来体检的53岁男性。这位医生与患者进行了交谈,并在患者的病历本中记录了结肠癌筛查、安全带、口腔护理、运动、改变膳食结构和使用防晒霜的重要性。同时也告知了患者前列腺癌筛查的利与弊。此后,他再也没有见过该患者。

碰巧的是,该患者又去看了第二位医生,这位医生没有与其讨论筛查的潜在利弊就直接安排了PSA筛查。PSA结果显示异常升高,很快又发现患者患上了不可治愈的晚期前列腺癌。虽然没有证据提示早期发现疾病就能改变患者的结局,但该患者还是起诉了第一位医生及其家庭医疗培训机构。

医生自己的话讲出了故事的另一半,“尽管我们掌握的所有国家权威机构的推荐意见均支持我这么做,并且文献也清楚地表明前列腺癌的筛查存在争议,但原告律师依然反驳……原告方认为我没有遵从英联邦弗吉尼亚标准行医。四位医生的证词是,当其接诊年龄大于50岁的男性患者时,他们不会与患者讨论有关前列腺筛查的问题,而是直接做检查。因为很可能超过50%的医生的确都是这样做的,因此,原告方的辩词极具说服力。可能会有人我们辩称,认为我们的做法比所谓的标准更合理,但这在法律上尚没有先例……开庭后7天,我被判无罪。我所在的住院医师培训机构却被判赔偿100万美元……在我看来,医学实践就应该紧跟当前可得的最佳证据并将其用于患者。在我看来,接诊患者时就应该采用共同决策的模式。在我看来,医生在实施检查时就应该将患者视为一个健康的人,而不是担心他将来是否会告你。我不确定我是否还会继续行医。”

Merenstein D. Winners and losers.

Journal of the American Medical Association,

2004; 291: 15-16.

上世纪80年代,日本首先开展了大规模的神经母细胞瘤筛查,但20年过去了,没有证据表明这种检测能降低神经母细胞瘤的死亡率。在没有无偏倚(公平)的临床试验证据证明筛查有益的情况下,日本就启动了这项筛查项目。而加拿大和德国纳入了3百万名儿童进行临床试验,结果显示筛查没有明显的益处,反而有明显的危害,这些危害包括不必要的手术和化疗,这两者都会产生严重的副作用。一位专家对加拿大和德国的研究结果直言不讳地评论道:

“神经母细胞瘤的筛查说明,人们很容易陷入“如果能够早期检测出某种疾病,筛查就一定有价值”的陷阱……这两项研究表明,神经母细胞瘤的筛查不但没有价值,还会导致“过度诊断”,也必定会查出那些可能自然消失的肿瘤。两项研究均提到筛查组儿童因接受治疗而出现了严重并发症。……希望将来在考虑开展其他筛查项目时能汲取这些

教训,如前列腺癌的筛查。”

前列腺癌与神经母细胞瘤不同,它是一种好发于成年男性的常见肿瘤(在英格兰和威尔士,前列腺癌是男性第二大常见肿瘤)。但筛查原则应该相同。前列腺癌的筛查是否符合筛查原则?血液中一种名为前列腺特异性抗原(PSA)的物质浓度升高会增加前列腺癌的死亡风险,但并无公开发表的无偏倚试验证明早期筛查能够改善患者的结局。然而可以肯定的是,PSA检测对人体有害。一些患者到了疾病终末期才接受治疗,此时治疗的意义并不大;而另一些患者因发现了本来并没有生命和健康危险的一类前列腺癌而接受了不必要的治疗。对于这两类患者,因PSA检测结果升高而开展的治疗可能导致令人痛苦的副作用,如大小便失禁或阳痿。

然而在美国和意大利,至少有1/3年龄大于50岁的健康男性接受过PSA检查。在美国赞成PSA筛查的呼声很高,这来自于公众、患者和医生。2001年,《旧金山年鉴》发表了一篇关于该市棒球队经理的文章。在常规PSA检测发现“阳性”结果后,这位经理刚刚做了前列腺癌手术。文章中对PAS筛查大加称赞;对于其弊端却只字未提。为了改变《年鉴》的男性读者对PAS筛查所持的完全乐观的态度,两位医生联系了报社,宣称该文没有反映出现实中围绕PSA筛查的广泛争论。于是他俩受邀写了一篇短文来讨论男性不应该做PSA筛查的原因。

#### 新生儿囊性纤维化筛查

- 利
- 为每个有囊性纤维化婴儿的家庭提供了获得专业治疗的机会;
  - 减少了因延迟诊断带来的悔痛;
  - 为将所有囊性纤维化病例纳入国家数据库提供了可能;
  - 有机会针对该病的治疗方法开展大规模的随机对照试验。
- 弊
- 尚无针对新生儿囊性纤维化筛查的完美试验,即现有检查项目可能存在漏诊,并且医生需要时刻警惕是否会在成年期诊断出该病;
  - 若诊断出基因携带状态,可能给患儿及其家属带来伤害;
  - 若不假思索和毫不考虑情感因素就将筛查结果告知患儿家属,也可能对家属造成伤害。

该文立即引发了强烈反响。在文章发表后数小时内,前列腺癌慈善机构、患者支持组织和泌尿科医生都给出了强烈的回应。撰文的两位医生遭到了来自Email的谩骂,将他们与纳粹医生Mengele相提并论,并谴责他们害死了成千上万的无辜者。这两位医生对此事引发的如此强烈的反响感到惊讶,他们写道:“原因之一在于支持PSA筛查的人对常规筛查有益于男性健康这一观点深信不疑。他们认为筛查的

确能‘改变世界’。我们对这种不切实际的想法提出了质疑,因而惹恼了这些人。我们也激怒了那些打着支持前列腺癌筛查的幌子,实为通过怂恿他人做检查而从中渔利的人。甚至一些患者支持组织也存在利益冲突,因为他们有赖于医药企业的支持。”

#### 筛查在发现囊性纤维化基因携带者方面存在的缺陷

虽然家属听到孩子没有患上囊性纤维化可能会松口气,但如果诊断为基因携带状态,随之而来的烦恼与悲伤可能会导致亲子关系受损、人格问题、关系失和、或类似于脆弱性儿童综合征的表现。除此之外,还可能引发亲子关系的否认(随后家庭破裂)、对小孩的歧视、购买医疗或人生保险困难,就业歧视(出于对携带者潜在危害的错误认识)和难以婚嫁。最后,如果婴儿的囊性纤维化基因突变不在标准囊性纤维化基因筛查试剂盒的检测之列,就有出现假阴性结果的危险。

David TJ. Newborn screening for cystic fibrosis.  
*Journal of the Royal Society of Medicine,*  
2004; 97: 209-10.

新生儿囊性纤维化的筛查又如何呢?这一致命性疾病通常在幼儿期就会出现症状,可导致慢性、消耗性胸部感染并最终发展成永久性肺损伤、食物吸收障碍、发育迟缓和肝脏衰竭。囊性纤维化是一种遗传性疾病,通常在儿童的一个基因中发生两处致病性突变时发病。若只有一处突变,则为疾病携带者,但自身无症状。除此以外,该病并不是这么简单,对有关囊性纤维化的遗传学基础了解得越多,该病就变得越复杂。现已发现还存在多种囊性纤维化的“非典型”类型。

近年来,随着物理治疗、抗生素和营养支持疗法的广泛应用,毋庸置疑,囊性纤维化患者的预期寿命已大大延长。理论上讲,通过筛查以实现早期诊断应该是有价值的,特别是在肺严重损伤之前。虽然目前对筛查试验的最佳组合尚未达成一致,部分国家已开始对新生儿进行筛查。

有利的一方面是,通过筛查实现早期诊断相对于出现症状后才诊断,患儿达到正常身高和体重的几率更大。但尚无法确定早期诊断是否能够防止肺损伤。并且,不应忽视筛查在发现囊性纤维化基因携带者方面存在的缺陷。如果携带者要自己生小孩,问题就会暴露;或者不用等到携带者自己生小孩,其兄弟姐妹就可能患上类似的疾病。正如两位研究者所言:“筛查提供了得到好结果的机会,但却不能确保得到的就一定是好结果。”

#### 筛查阻生智齿是明智的做法吗?

常规口腔检查是最常见的筛查项目之一。然而数年前就有证据表明这种筛查弊大于利。原因之一是口腔筛查可能导致智齿被拔除。智齿是最后长出的恒牙,出牙时间通常在18~24岁之间。但有时由于各种原因,它们未能从牙

龈长出,即阻生。通常阻生智齿不会引起任何问题,但有些人的确会并发周围牙龈感染,并累及周围的牙和骨。虽然毫无疑问应该拔除有问题的阻生智齿,但拔除健康的阻生智齿则另当别论。此外,拔除这些智齿既痛苦又昂贵——仅在英格兰和威尔士,NHS 在这类牙科手术上的花费就达数百万英镑。因此,负责基于证据发布推荐意见的英国国立临床卫生优选研究所(NICE)受命于 NHS,对此问题进行调查并给出指导意见。在评价了相关证据后,NICE 于 2000 年直截了当地公布了研究结论:没有致病的阻生智齿不应被拔除。原因有二:(1) 尚无可信的研究证明拔除智齿对患者有益;(2) 被拔除了健康智齿的患者承担着手术风险。这些风险包括神经损伤、其他牙齿损伤、感染、出血和死亡(罕见)。并且,患者术后可能出现肿胀和疼痛,以致于无法完全张开嘴。

### 卧床休息就一定正确吗?

我们已经了解了过度治疗和过度热衷于筛查可能弊大于利,然而甚至象卧床休息这样看似无害的建议也可能产生误导。人们通常认为卧床休息对多数疾病都有好处。然而,如果把卧床休息作为一种加速病情恢复的治疗措施(已经用于多种疾病和术后患者),与其他治疗措施一样,也应该严格评价其利弊。上世纪 40 年代,以术后患者为研究对象的试验表明完全卧床休息不但无效,反而会带来诸如小腿血栓和褥疮等潜在危险,至此卧床休息的价值首次遭到了质疑。那么,关于卧床休息作为一种治疗措施的利与弊的无偏倚证据是什么? 1999 年,澳大利亚的研究者决定对已发表的关于卧床休息作为一种治疗措施的无偏倚(公平)研究进行系统

评价,以找到有关其利弊的证据。研究者共找到了涉及 15 种疾病的 39 个临床试验,共纳入患者近 6000 例,对治疗可能产生的多种效应(有益效应和有害效应)进行了评价。

卧床休息作为一种治疗措施主要有两大用途:一是作为内科或外科操作后的预防措施,二是作为一线(主要)治疗。24 项对照试验表明,术后卧床休息未见明显益处。9 项试验显示,部分操作后卧床休息反而有害,包括腰穿和腰麻后卧床休息。15 项试验提示,卧床休息作为不同疾病的一线治疗未见明显益处。还有 9 项试验得出了卧床休息对某些疾病有害的证据,包括腰背痛,分娩和心脏病发作。总之,对于研究所涉及的疾病,疗效证据显示卧床休息可能延迟病情的恢复,甚至对人体有害。

#### 卧床的危险

“生活中我们向来惧怕  
把时间无端地浪费在床上。  
把患者从床上唤起  
将免于其过早地步入坟墓。”

*Asher R. The dangers of going to bed.*

*British Medical Journal, 1947; 14 Dec.*

*Reproduced in: Jones FA, ed. Richard Asher talking sense.*

*London: Pitman Medical.1972.*

#### 要点

- 高强度的治疗并不一定有效;
- 对健康人群进行疾病筛查可能弊大于利。

## 第3章 公平试验的主要概念

前两章介绍了未充分验证的治疗措施可能造成严重危害。显然,严格评价治疗措施对于确定是否应将治疗措施用于患者是非常重要的。

误导人的宣传随处可见,因此我们每个人都需要有能力去判断这些有关疗效的宣传是否真实可信。如果不具备这样的能力,我们很可能把实际上无效的治疗措施认定为有效,而把实际上有效的治疗措施认定为无效。为了公平地验证治疗措施,我们必须致力于获取可靠的疗效信息。而其中最重要的一点是,必须减少偏倚和机遇因素对疗效的歪曲。那么怎样才能做到这一点呢?

### 推翻谬论

当 James Lind (参见第1章)开始查阅有关坏血病的文献时,他发现仅有的疾病描述居然都来自于没有受过专业训练的海员和从来没有出过海的医生。“没有一位熟悉这种疾病的医生致力于相关治疗的研究。”Lind 认为这就是造成该病的诊断、预防和治疗一直疑问重重的原因之一。Lind 坦率地写道:“在这个问题被正视和阐明之前,有必要先推翻大量的谬论。”

*Lind J. A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with critical and chronological view of what has been published on the subject.*

*Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson, 1753, pviii.*

### 理解偏倚

验证治疗措施的过程中出现的偏倚是指那些导致疗效结论与事实不符,产生了系统误差而不仅仅是偶然误差的影响因素。虽然有多种偏倚都可歪曲卫生保健研究的结果,但在验证治疗措施的公平试验中必须减少的偏倚主要有:

- 受试人群的差异导致的偏倚
- 疗效评价方式的差异导致的偏倚
- 可获取证据的报告偏倚
- 可获取证据的选择偏倚

这里所涉及的原则对许多读者来说是较为陌生的。部分读者会觉得本章是全书最具挑战性的一部分。主要问题的完整信息和解释可参照 James Lind 图书馆([www.jameslindlibrary.org](http://www.jameslindlibrary.org)),希望这会有助于读者理解本章内容。

### 验证治疗措施的公平试验

对照是所有公平试验的重点;对于判断某种治疗是否

能够产生某种效应必不可少。有时是将2种或2种以上的治疗做比较,有时也可将一种治疗与非阳性治疗做比较。无论采用哪种对照,都应解决真实存在的有关疗效的不确定性,这种不确定性是指没有基于研究的确切证据(第1章介绍了不确定性的概念,第4章将介绍相应的处理方法)。要实现真正意义上的公平对照,就必须尽可能地减少偏倚。

### 为什么对照必不可少

只要思考片刻,就很容易理解治疗对照的必要性。“大自然就是最好的医生”这句古老的谚语说得好,人们常常没有接受任何特殊治疗,病自然就好了。因此在验证治疗措施时,必须考虑疾病的自然病程以及不接受治疗的情况下疾病的结局。治疗可能对疾病的自然结局产生积极或消极的影响,也可能完全与疾病的结局无关。

### 与我无关

“当然,这次患者得以痊愈与其接受的治疗无关,与我的医术也无关。从另一方面来讲,这也清楚地证明相对于人体自身的恢复力而言,治疗措施显得并不那么重要。”

*Cochrane A. Sickness in Salonica: my first, worst, and most successful clinical trial.*

*British Medical Journal 1984; 289: 1726-7.*

包括医生和患者在内的所有人有时都会主观地比较治疗措施的疗效。这往往会形成一种印象:新疗法的疗效优于旧疗法。然而这些主观的判断都必须通过正规的研究来验证,例如最初可通过分析临床记录。这样的验证可能会要求人们对新旧疗法进行谨慎的对照。

如果单凭主观印象去指导临床实践,就会产生风险(参见第1章)。基于主观印象或初步分析的治疗对照往往是不可靠的。只有当治疗效果非常明显时,这种对照才有意义,如阿片用于镇痛、胰岛素用于治疗糖尿病、髋关节置换术用于治疗骨关节炎。然而在多数情况下,治疗效果不会这么明显,这就需要谨慎地避免有偏倚的对照和错误的结论。

将当前施行的治疗与过去施行的治疗进行对照通常是不可靠的,因为其他相关因素也在随时间而改变。第1章中举过一个例子,激素己烯雌酚(DES)预防复发性死产,这个例子就清楚地证明了这一点。初产妇比经产妇更容易出现死产,因此,将使用了DES的经产妇的死产率与没有使用DES的初产妇的死产率进行比较,得出的结果是DES可降低死产的风险,这一结论具有严重的误导性。这个例子同时还指出,孕妇接受DES治疗可能会给她们的下一代带来灾难性的后果。因此,应尽可能在同一时间点下对不同的治疗措施进行比较。

### 为什么对照必须解决真实存在的不确定性

在从事新的治疗验证试验之前,有必要先了解已有的研究结果。虽然这看起来是理所当然的,但由于人们往往忽视了已有的可靠证据,而产生了关于治疗效果的不确定性。应该系统、严格地评价证据,确信提出的新的治疗验证试验将解决当前存在的不确定性。如果一开始就忽视了这一关键步骤,后果将非常严重,患者可能遭受不必要的痛苦,宝贵的卫生资源和科研经费可能会被浪费。这一切是怎么发生的呢?

20世纪90年代初期,美国研究团队从医学教材和期刊找出了30年来有关心脏病治疗措施的推荐意见,并将这些推荐意见与本应该系统评价的公平试验结果形成的证据进行了比较。研究者发现,由于编写教材的作者评价证据时,没有尽可能地减少因偏倚产生的误导结果和机遇因素的影响,而给患者造成了严重后果。有的患者被剥夺了获取那些可挽救生命的治疗建议的机会(如溶栓药治疗心脏病发作),有的甚至长达10年之久;而有的医生一直在推荐那些早已被公平试验证实为有害的治疗措施(如抗心律失常药物治疗心脏病发作——参见第1章)。

那些没有回顾过去的试验就开展新试验的研究者可能不会意识到,他们所要验证的治疗效果的不确定性可能已经得到了充分证实。这就意味着一些患者参加了不必要的试验研究,部分患者还失去了接受有效治疗的机会。例如,可靠证据显示,对接受肠道手术的患者给予抗生素可降低手术并发症导致的死亡率,然而在之后很长一段时间里,仍有研究者继续就此开展对照试验,让一半的受试者不接受抗生素治疗。反之,如果评价了以前的研究结果后明显发现可靠证据缺失,此时就肯定需要开展新的试验。

如第1章所述,如果研究者在开展人体试验之前没有系统地评价来自于动物研究的相关证据,患者也可能因此受到伤害。例如,如果先评价了动物实验的结果,就绝不会在中风患者中开展尼莫地平的临床试验了。

系统地评价研究证据,这决不是一个新观点。早在1753年,James Lind所著的《坏血病大全》一书中(书中报告了他所开展的公平试验,后来该试验结果被广泛采纳)就蕴含了“按时间先后顺序严格评价已发表的相关证据”这一观点。

### 避免有偏倚的对照

为了确保公平的对照,必须识别出并尽可能减少各种偏倚。如果做不到这一点,一项实际上并不比现有治疗更有效的新治疗就可能表现出虚假的优势。

在考虑单项试验研究时,偏倚可能来源于:

- 将病情相对较轻又接受了新治疗的患者与病情相对严重却接受了标准治疗的患者进行对照;
- 对治疗结果的评价有偏倚,例如将知道用的是价格昂贵的新治疗并且认为新治疗更有效的医生和患者的意见与那些知道用的是已有标准治疗的患者和医生的意见进行对照。

在评价多个相似研究时,偏倚可能来源于:

- 仅纳入那些说明新的治疗更有效的试验,而不考虑其他得出了“阴性”结果的试验,即未能证实新治疗的优势或提示新治疗可能有害(“阴性”结果通常都没有报道);

- 对可获得证据的选择和诠释存在偏倚。

人们往往没有意识到这些不公平试验的结果存在偏倚,就轻易地作出了治疗决策。然而,令人担忧的是,某些获取了利益的人有时会竭力掩盖偏倚,使治疗措施看上去比实际更有效。一些研究者常常但并不总是出于商业目的,故意忽略一些已经存在的证据。他们设计、分析并报告研究,以此来粉饰某项治疗措施的疗效。

### 利用偏倚

“由药厂赞助的研究比起那些由其他资金资助的研究更可能得出有利于赞助商产品的结果。在过去至少20年内,这些结果被广泛用于各种疾病状态、药品和药品分类,却没有人去评价其研究类型。”

*Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review.*

*British Medical Journal, 2003; 326: 1167-70.*

### 因受试者差异产生的偏倚

病情相对较轻的一组患者接受一种治疗,而病情相对较重的一组患者接受另一种治疗,那么这两种治疗的对照就是不公平的。有时,这个问题可以通过交叉试验得以解决,即在不同的时间给予同一组患者不同的治疗以此来进行对照。但在很多情况下这种试验设计是不可行的,例如,不可能用这种设计来对照不同的手术治疗。

经常都是通过比较接受了不同治疗的各组患者来评价治疗措施。要做到公平地比较,那么各组患者必须相似,这样才能实现同类比较。如果接受某治疗措施的患者相对于接受另一种治疗措施的患者更可能好转或恶化,这样的偏倚就使得即便得出了有差异的结果,也不可能真实地反映治疗的效果。18世纪,外科医生Willian Cheselden意识到了这个问题。当时,医生正热衷于比较膀胱结石术后患者的病死率。但Cheselden指出老年患者死亡的可能性更大,因此在比较由不同外科医生实施的各种类型的手术后患者的病死率时,应该考虑手术患者的年龄差异。

比较那些过去曾接受了不同治疗的患者的经历和治疗结局,这在今天仍然是评价疗效的方式之一。关键是要弄清楚在接受治疗之前,各组患者的情况是否足够相似。例如,通过比较使用了激素替代疗法(HRT)与没有使用HRT的妇女的发病率来评价HRT的效果,这是一种非常危险的极具误导性的评价方法。这种对照提示HRT能够降低心脏病发作和中风的风险,但后来的随机试验显示HRT产生的结果恰好相反(参见第1章)。因此,没有考虑这类偏倚的研

究不仅毫无用处,而且危害了患者。

最好的办法是在开始治疗之前就设立对照。1747年,James Lind在HMS Salisbury上对6种坏血病治疗措施进行对照之前,他特意选择了那些处于这种致命性疾病同一阶段的患者(参见第1章)。他还保证患者有相同的饮食和住宿条件。这表明他清楚地意识到了,这些非治疗因素而非治疗措施本身可能影响船员康复的机会。

这正如在今天我们必须谨慎确保各治疗组患者的基线情况是类似的。只有一种方法可以做到这一点:必须采用某种基于概率的方法来分组,即“随机分配”,这是“随机”公平试验唯一的且非常重要的特征。

抽签方法,如掷骰子,将确保各治疗组患者的基线情况相似,基线情况不仅包括已知且已测量的重要因素,如年龄,还包括可能影响疾病康复的非测量因素,如饮食、职业和其他社会因素,以及对疾病或治疗的担忧程度。避免分组偏倚的最佳方法是确保患者和医生都不知道患者的分组情况。

通过这种方式分组实现了同类比较之后,还应避免因忽略了部分患者的变动情况而导致的偏倚。这就意味着,无论患者实际接受了何种治疗,都应尽可能地对所有接受了治疗分组的患者进行随访,并且都要纳入所属治疗组的结果的主要分析,这就是所谓的“意向治疗”分析。

这种方法似乎看起来不合逻辑,但如果忽略了这类患者可能导致整个试验不公平。以有缺血性中风危险的患者为例。研究者开展一项试验以验证血管溶栓术是否能降低那些因血管栓塞常出现晕眩的患者中风的风险,将手术组患者与非手术组患者进行比较。如果他们仅记录那些手术后存活下来的患者的中风发生率,该试验就忽视了一个重要的事实:手术本身就可能引起中风和死亡。因此,对手术效果的评价就是不公平的。

### 随机的原因

“参与随机试验的医生不应将自己视为一名研究者,而就是一名对患者负有伦理责任的临床医生,在没有尽全力评价治疗措施的真实效果之前,绝不将其用于患者。”

Rees G, ed. *The friendly professional. Selected writings of Thurstan Brewin.*

Bongnor Regis: Eurocommunica, 1996.

### 治疗结局的评价偏倚

大多数患者和医生都希望治疗措施是有效的。这种乐观的态度对于患者就治疗的满意度有着非常积极的作用,正如英国医生Richard Asher在一篇他为医生所著的文章中写道:

“即便对照试验显示某治疗毫无用处,但只要你坚信你所开展的治疗,治疗结果就会更好,你的患者就会更好,你的

收入也将更好。我认为这正是那些天资不高但容易轻信他人的医生却取得了巨大成功的原因,也正是许多医生厌恶其他时髦又成功的医生所喜欢展示的统计学数据和对照试验的原因。”

即使医生知道自己开出的治疗是没有任何“生理”作用的,他们还是会开,希望能通过心理作用来帮助患者。换言之,即便实际上用的是模拟药物(安慰剂),只要患者相信它能够缓解症状,就可能明显感受到病情改善。

因此,在开展验证治疗措施的公平试验时,必需减少医生和患者评价治疗结果时可能出现的偏倚。为此人们常采用一种被称为“盲法”的技术。“盲法”的兴起还有着一段有趣的历史。18世纪,法国国王路易十六下令研究弗朗兹·麦斯麦提出的动物磁力说(麦斯麦术)。国王想要知道这种效果是归功于某种“真正”的力量或“精神幻觉”。被蒙上了双眼的受试者被告知其接受了或没有接受麦斯麦术,事实上情况是完全相反的。结果是,只有那些被告知接受了麦斯麦术但实际上并没有接受的受试者才感受到了效果。

对于某些结局,如死亡,出现评价偏倚的可能性极小,因为几乎没有余地去怀疑一个人是否真的死了。然而,对大多数结局的评价都带有一定的主观性,如患者的症状。人们可能出于个人原因而偏爱某种治疗:如果患者相信某种治疗对他们是有利的,就可能更多地留意提示疗效的征象;相反,如果患者对某项治疗存在疑虑,就可能更急于将某些负面反应归咎于这种治疗。

在这种情况下,就需要通过盲法来实现公平评价。所比较的两种治疗必须在外观上看起来一致。有时采用的是安慰剂,这是一种没有任何生理活性(模拟)的处理措施。例如,20世纪四五十年代英国医学研究委员首先对普通感冒的治疗进行了评价,如果没有采用外观相似的安慰剂来防止患者和医生知道患者接受的是新药或安慰剂,那么将难以解释试验的结果。这就是所谓的双盲试验。

下面再举一个例子来说明双盲法的重要性。研究者评价了对医生设盲(医生不知道患者接受的是药物还是安慰剂)对于一项多发性硬化临床试验结果的影响。在试验期间的每一次评价中,设盲的医生和没有设盲的医生都对所有患者进行检查;每位医生都对结果进行评分。没有设盲的医生的评分结果显示某一项治疗措施优势显著,但设盲的医生却没有得出这一结果。但采用设盲的医生的评分结果就避免了得出错误的结论。综上所述,结局指标的评价主观性成分越高,就越需要采用盲法来保证试验的公平。

尽管,有时对医生和患者进行盲法处理是不可行的,例如很难掩盖手术治疗与药物治疗之间的差异。一些明确的结局,如死亡,就没有可能出现评价偏倚。即便当偏倚不知不觉地发生时,例如在评价死亡原因时,仍可以让不知道患者接受了何种治疗的医生来寻找死亡原因。

### 如何解释无偏倚对照

### 考虑预期治疗与实际接受治疗之间的差异

就本章目前所给出的所有原因而言,你已经意识到公平试验必须仔细规划。制定这些规划的文件称为试验方案。试验方案明确说明了所要对照的治疗措施的细节。然而,计划得再好,不一定都能落实,即患者实际接受的治疗有时会与应当接受的治疗不同。例如,患者可能不会接受预期的治疗;或某种治疗措施无法实施。一旦出现了这类情况,人们在分析和阐释结果时就应当仔细思考并加以说明。

### 考虑机遇因素

对照两种治疗措施时,治疗结果的差异可能就是机遇因素的作用。如果5名患者在一种新疗法的治疗下病情好转,而7名患者在标准疗法的治疗下病情也好转,此时没有人能够肯定地说新疗法不如标准疗法。如果再进行一次对照,病情好转的患者人数可能会与原先相反(7比5),或相同(6比6),或出现其它比率。

然而,如果50名患者在新疗法的治疗下病情好转,而70名患者在标准疗法的治疗下病情好转,就不大可能再用机遇因素来解释这种差异了。若这个数字变成了500人和700人,那么很明显,新疗法的确不如标准疗法(实际上大约半数的新疗法都是如此)。

因此,要减少治疗验证试验中机遇因素对患者的误导,就必须从人数足够多的患者中得出结论,无论这些患者病情好转或恶化,还是既无好转也无恶化。

为了评价机遇因素对公平试验结果的影响,研究者对“统计学意义”进行了检验。这种检验有助于避免得出错误的结论,即当治疗措施之间并不存在差异时认为彼此有差异,或当治疗措施之间真正存在差异时认为其没有差异(这种情况更常见,也更危险)。当研究者和统计学家讨论不同治疗之间的明显差异时,他们指的是统计学意义。但是要记住,统计学显著性差异并不一定提示差异是“显著的”。不同治疗之间的差异排除机遇原因外,即使有统计学显著性差异,也可能没有临床意义。例如,一项随机试验的系统评价把成千上万每天吃1粒阿司匹林的健康人与成千上万不服用阿司匹林的人相对照,发现服用阿司匹林的人患心脏病的几率更低。这一差异具有统计学意义。那就是说,这是不可能用机遇因素来加以解释的。然而,这一发现并不一定具有实际意义。如果一个健康人患心脏病的风险已经很低了,那么通过服药来进一步降低患病率是不对的,尤其是阿司匹林还具有一定的副作用。

有一种办法可以降低患者被机遇因素误导的可能性,那就是估算可信区间。可信区间给出了治疗措施实际效应量(永远都无法精确得知)的范围大小,并且反映了肯定的程度(通常是95%或99%)。这就像是问你问别人:“你上班路上要花多长时间?”别人回答说:“二十分钟到一小时不等,要看交通状况。”

因此,统计学检验有助于评价机遇因素,避免在治疗措施之间不存在差异时却得出有差异的结论,或有差异时却得

出无差异的结论。

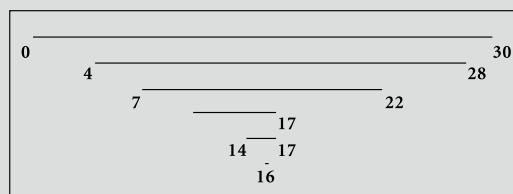
### 可信区间

如果口袋里有30颗糖(看不到里面的内容),可能是橘黄色,也可能是白色,那么一开始口袋里橘黄色糖的数量应该介于0到30之间。橘黄色糖数量的可信区间是0-30。

如果从口袋里抓出一把糖,发现手里有4颗橘黄色的和2颗白色的,那么你就知道一开始至少有4颗橘黄色的糖,但最多不超过28颗。也就是说,在最初的30颗糖里至少有2颗是白色的。因此一开始橘黄色糖数量的可信区间是4-28。

如果再抓出一把糖(不与第一次重复),其中有3颗橘黄色的和6颗白色的,那么一开始口袋里橘黄色糖数量的可信区间就是7-22。如果再抓一把,有3颗橘黄色的和5颗白色的,那么一开始橘黄色糖数量的可信区间就是10-17。又抓一把,有4颗橘黄色的,那么可信区间就是14-17。抓出最后的3颗糖,有2颗橘黄色的和1颗白色的。至此可以得出结论,一开始共有16颗橘黄色的糖。抓出的糖越多,可信区间就越窄。这个例子计算出了完整可信区间——每一步我们都可以完全确定橘黄色糖的实际数量介于可信区间两端之间的某一点。

下图阐释了上述例子,反映了可信区间是如何一步步变窄的。



Adapted from Critical Appraisal Skills Programme, Cochrane Consumer (UK). Consumers commenting on Cochrane Reviews. Post-workshop

摘自 Critical Appraisal Skills Programme, Cochrane Consumer (UK). Consumers commenting on Cochrane Reviews. Post-workshop pack. 2003-4, p23.

### 发现并评价治疗措施的非预期效应

治疗措施的初始评价,如那些正在申请上市的新药,最多可涉及几百人或几千人,治疗时间一般就是几个月。在这个阶段中,可能获得相对短期且出现频繁的非预期效应,而那些罕见的、需要花费时日才能产生的效应只有等到该治疗措施得以更广泛应用时才能发现。例如,当医生按惯例开了一种新药,患者可能与那些被选来参加临床试验的人有所不同:他们可能更年长或年轻、性别不同、病情不同,或除了治疗所针对的疾病外还有其它健康问题。非预期效应,无论好坏,常常是由医务人员或患者最先发现。可是在这些预感中

有哪些反映出了真正的效应呢?

如果非预期效应非同寻常,并且在治疗后经常发生,则通常会令医生和患者觉得是不是有什么地方出了问题。沙利度胺(反应停)就是这种情况(参见第1章):婴儿出生时没有四肢,这几乎是先前没有听说过的。同样,有时也会发现非预期的有益效应。例如,发现用来治疗精神分裂症的药物也可以降低胆固醇。如果能注意到这些明显的关系,通常都能够确认其成为真正的非预期效应。

然而,许多关于非预期效应的预感都是建立在并不十分令人信服的证据之上。与设计用于发现预期效应的试验一样,用于证实或驳回并不显著的疑似非预期效应的试验也应该避免有偏倚的对照。这类试验就必须遵循“同类比较”原则。

有时,研究者能够对那些参与早期试验的人进行进一步的分析或随访,试验中通过随机分配实现了治疗组之间的可比性。通常这是无法选择的。因此,组建新的无偏倚的对照组更具挑战性。在这里,效应都是非预期的。非预期效应通常是不同于治疗所针对的疾病或病情的另一种疾病或病情。例如,荷尔蒙替代疗法(HRT)最初是用来缓解更年期症状的,而妇女患乳腺癌的风险是没有被考虑在内的。换言之,没有明显的理由认为接受了HRT的妇女与没有接受HRT的妇女相比,患乳腺癌的风险不同。这就是证实了HRT会增加乳腺癌风险的公平试验的基础。

当疑似非预期效应是由诸如心脏病一类的常见病的治疗所致,但在新疗法中却并不常见,此时就只有通过调查大量接受了治疗的患者才能发现非预期效应。例如,20世纪60年代末,尽管一些研究者认为阿司匹林可能会降低患心脏病的风险并开始进行公平试验,但多数医生都认为这是完全不可能的。一项旨在发现药物的非预期效应的大规模试验表明,因心脏病入院的患者与同类患者相比,近期服用过阿司匹林的患者比例很小,这时医生们才开始转变观念。

20世纪70年代末,在经历了沙利度胺灾难之后,人们首次确立了发现并评价非预期效应的基本原则。从那时起,随着许多强有力治疗措施的问世,这类针对非预期效应的公平试验至今仍极具挑战性和重要性。

### 考虑所有相关证据

统计学家 Austin Bradford Hill 是公平验证治疗措施的先驱之一,他曾经说道:“阅读研究报告的人都希望能够解答以下4个问题:为什么开展研究?研究具体做了什么?发现了什么?研究结果意味着什么?”对最后一个问题的回答尤为重要,因为这影响到治疗措施和未来研究的选择与决策。仅靠一项公平试验很难提供足够强大的证据来给出可靠的回答。通常需要开展多项公平试验来解决同一个问题。因此,为了回答“研究结果意味着什么?”,我们必须把来自一项试验的证据与来自其它试验(都是为了解决同一个或类似的问题)的证据结合起来进行阐释。

一百多年前,英国科学进步协会的主席 Lord Rayleigh 曾经就这一原则发表评论:

“正如人们有时所预料的,如果科学仅仅是由事实不断累积而成,那么它很快就会陷于停滞并被自己摧毁……获取新知识、消化吸收旧知识,这两个过程是同时进行的。因为两者都必不可少,所以我们可以讨论它们的相对重要性。然而,我们必须做出评论。发现问题与解释问题应该是并驾齐驱的,不仅要陈述新的事实,还应指出新事实与旧事实之间的关系。这些工作是值得我们去做的,但很遗憾,人们通常都做不到这一点。”

然而,即便在今天, Rayleigh 的明智建议也常常被忽视,因此导致读者在阅读新的研究报告时,很难获取“研究结果意味着什么?”这一问题的可靠回答。此外,在报告试验结果时如果不考虑其他经过了系统评价的相关证据,就会推迟识别出有益与有害治疗的时间。例如,20世纪60年代至90年代初,研究者开展了50多项公平试验以评价药物是否能降低心脏病患者心律失常的发生率,之后才意识到这些药物事实上正在扼杀人的生命(参见第1章)。如果每份研究报告都能参照其它相关证据对新的结果进行评价,那么早在10年前就会发现这些药物的致命性副作用。

### 处理可获取证据的报告偏倚

解释新的研究结果时应当先对其它所有可靠的相关证据进行系统评价。这句话听起来很容易,但从许多方面来看都是一种挑战,一来是因为一些相关的证据没有发表,二来那些得出了令人失望的或负面的结果的研究很少被报告。这种“报告偏倚”主要源于研究者没有撰写或提交研究报告用以发表。但有时候,杂志社在退稿时也存在偏倚。还有一个问题,研究者可能会选择性地掩盖那些与疗效解释相抵触的结果。

为了减少报告偏倚,所有验证治疗措施的公平试验一开始就需要注册(参见 [www.controlled-trials.com](http://www.controlled-trials.com))。最重要的是,所有的临床试验结果都应该发表,无论其是否会令研究的申办者或研究者本人感到失望。对研究报告不足,导致报告偏倚,这是一种不科学、不道德的行为。近来,一些掩盖药效负面证据的事件曝光后,激起了强烈的公愤并且有人开始起诉药厂。这些事件终于令人们长期以来的需求得以满足,那就是要求临床试验从一开始就公开注册,并且发表所有的试验结果。在电子出版物问世以前,要强制执行这些原则是很困难的,但随着电子期刊的开放获取,例如 BioMed Central ([www.biomedcentral.com](http://www.biomedcentral.com)) 和公共科学图书馆 ([www.plos.org](http://www.plos.org)) 发表的期刊,这个障碍已经逾越。

### 避免可获取证据的选择偏倚

偏倚不仅会歪曲单项试验的结果导致得出错误的结论,还会歪曲对证据的评价。评价是很重要的,因为大多数人都依赖于评价,但是评价必须是系统的,否则就会误导他人。

例如,评价者可能只对自己熟悉的研究进行了评价,那么所得结论就可能存在偏倚。

为了避免这些问题,应当制定系统评价计划书,明确将要采取何种措施来减少偏倚。这些措施可能包括:阐明这次评价旨在解决哪些问题;所评价研究的纳入标准;如何发现可能合格的研究;筛选研究时将采取哪些措施来避免偏倚。

那些看似为了回答同一个问题的系统评价往往会得出不同的结论。有时这是因为所要回答的问题存在细微的差异,有时这又反映出评价者所采用的方法存在差异。在这些情况下,就有必要判断哪种评价最有可能成功地降低了偏倚和机遇因素的影响。

是否还存在其他利益冲突会影响到研究者对评价的实施或解释?的确有,比如研究者与生产新药的药厂有着某种直接关联。在评价治疗湿疹的月见草油的疗效时,那些与药厂有关联的评价者得出的结论比那些没有这种商业利益的评价者得出的结论更乐观(见第1章)。

商业利益并不是导致证据选择偏倚的唯一因素。无论研究者、医务人员还是患者,都可能因偏见而导致选择偏倚。

### 采用Meta分析减少机遇因素

为了减少机遇因素,有时可以将所有相关研究的结果进行统计学合并,这称为Meta分析。尽管多年以前统计学家

就制定了Meta分析的方法,但直到20世纪70年代才开始得以广泛应用,由最初的社会科学家到后来的医学研究者,都在使用这种方法。到20世纪末,人们已广泛认可Meta分析是治疗措施公平试验的重要组成部分。

Meta分析是另外一种有助于避免错误结论的方法,即当某种治疗实际上有益或有害时,结论却是该治疗没有任何效果。例如,可能出现早产的孕妇是否应该短期使用价格低廉的类固醇。1972年报告了第一项随机对照试验,该试验表明这种治疗可降低婴儿的死亡率。十年之后,出现了更多的试验,但都是小规模,不同的试验其结果也不尽相同。那时,还没有开展系统评价通过Meta分析的方法来合并证据。如果当时就开展了Meta分析,结果将有力地支持类固醇的疗效。然而直到1989年,才首次发表运用了Meta分析的系统评价。因此,大多数产科医生都没有意识到该治疗的疗效,从而导致了数万名早产婴儿患病或死亡,而这一切原本都是可以避免的。

### 要点

- 如果不重视偏倚和机遇因素的影响,就很容易得出错误的结论,将实际无效的治疗视为有效,将实际有效的治疗视为无效。
- 对照是所有公平试验的关键。
- 选择性地报告试验结果将危害患者。
- 系统评价所有相关证据应当作为疗效评价的基础。

## 第4章 应对治疗措施效果的不确定性

第3章概述了应如何公平验证治疗措施。本章将深入探讨几乎总是存在的有关治疗措施效果的不确定性。

20世纪70年代,本书作者之一(IC)在美国度假时扭伤了脚踝,一位骨科医生对其进行了诊治。这位医生将患足用夹板暂时固定,建议在肿胀消除后用石膏模型固定小腿6周。回国数日后,IC前往当地的骨科诊所就诊,一位英国骨科医生毫不犹豫地否定了这个建议。这位医生说,将患足用石膏固定是完全不恰当的。鉴于这种显然不确定哪种措施更好的情况,IC要求接受对照试验以找出答案。这位英国医生却回答到,对照试验是用于那些不确定自己的判断是否正确的人,而他确信自己是正确的。

专家意见为何会出现如此大的分歧?患者又该如何选择?每位医生都坚信自己采取的治疗措施是完全正确的。然而,医生们的观点却存在明显差异,这清楚地暴露了整个业界对于普通骨折的最佳治疗方案尚存在不确定性。有高质量证据证明哪种治疗措施更好吗?如果有,是哪一位甚至这两位医生都不知晓吗?或者根本就没有人知道哪种治疗措施更好?也可能是这两位医生对治疗的某些结局持有不同的价值观:美国医生可能更关注疼痛是否能缓解,所以建议用石膏固定,而英国医生可能更担心是否会出现肌肉萎缩(肢体长期制动后易出现)。如果真是这样,为什么两位医生都没有询问IC哪种结局对他这位患者更重要?

### 医学不确定性的演进

我的老师 Tacoltt Parsons 最先让我意识到了现代医学实践中诸多不确定性的重要,他将其视为一种理论概念、经验现象和人类经验。他也告知我医生和患者都存在一些严重的认识误区,20世纪医学科学技术领域的伟大进步让我们看清了在探询健康与疾病、生与死的道路上我们仍然是多么无知、迷茫,还在犯着多少错误。

*Fox. R. The evolution of medical uncertainty.  
Milbank Fund Quarterly. 1980; 58: 1-49.*

这涉及到多个方面的问题。首先,有比较这两种迥然不同的治疗方案的可靠证据吗?如果有,这些证据能够反映IC或其他患者可能关注的治疗结局(如缓解疼痛或避免肌肉萎缩)吗?不同的患者关注的治疗结局可能有所不同。如果没有能够提供所需信息的证据,又该怎么办?

一些临床医生很清楚在没有可靠的证据时应该如何处理。例如,一位专攻脑卒中治疗的医生这样说道:“我可以向患者保证,我是脑卒中评估和诊断方面的专家,我能够正确解读脑扫描结果并施行正确的检查。我也从现有研究中了解到,我的患者在卒中单元接受治疗后恢复得更好。尽管如此,我以及其他医生仍有一个问题不能确定,即是否应该使用抗血栓药物,这些药可能利大于弊,但在实际运用中可能弊大于

利。在这种情况下,我认为应该向患者解释我只准备在有严格对照试验的前提下才使用这种药物,以此来减少不确定性,这是我的职责所在。”本章重点讨论当备选治疗措施的疗效信息不足,患者也没有明显的偏好时,应如何应对这种不确定性。

### 疗效显著:罕见且容易识别

有关治疗措施效果的不确定性几乎是不可避免的,因为只有极少数的治疗措施能够发挥出令人毋庸置疑的疗效。此时,治疗措施的效果非常明显。例如,20世纪30年代发现的磺胺类抗生素治疗当时妇女分娩时常见的致命性疾病——产褥热。产褥热是因产道细菌感染所致,常见致病菌是化脓性链球菌。19世纪70年代末,尽管引入了严格的抗感染措施,全球因产褥热丧命的妇女仍高达数千人。磺胺类药物的使用起到了显著的效果,病死率大幅下降。磺胺类药物治疗重症细菌性脑膜炎(脑膜炎)也称脑膜炎球菌性脑膜炎也有显著的疗效,同样大幅降低了病死率,所以这样的疗效是不容怀疑的。在这些情况下,就无需再开展严格的对照试验来证明磺胺类药物的疗效,因为只需对照过去没有使用磺胺的患者,就会发现磺胺的疗效是非常明显的。

第3章也提到了两个疗效显著的例子:阿片镇痛和胰岛素治疗糖尿病。例如,20世纪20年代,加拿大医生 Banting 和 Best 发现了胰岛素(胰腺产生的激素),在这之前糖尿病患者的寿命短暂且受尽折磨,常因血糖控制不佳而极度消瘦。动物实验的初步结果公布后,很快就将胰岛素用于患者,治疗取得了巨大的成功,当时胰岛素的疗效近乎神奇。

发生于同一年代的还有肝脏制品治疗恶性贫血(后来发现肝脏制品中含有维生素B<sub>12</sub>)。在当时这被视为一种致命性贫血疾病,红细胞会逐渐降到非常危险的极低水平,患者极度苍白和虚弱。采用肝脏提取物后,不但起效快而且效果显著。如今维生素B<sub>12</sub>已是该病的常规用药。同样,20世纪40年代,链霉素治疗结核性脑膜炎、青霉素治疗各种细菌性感染的疗效也是显而易见的。

近年来,器官移植用于肾脏、肝脏或心脏衰竭患者,髋关节置换术用于关节痛患者,这些治疗措施的疗效都是十分显著的,无需再进行严格的对照试验。本世纪初,伊马替尼治疗慢性粒细胞白血病也取得了类似的显著疗效。在伊马替尼问世之前,针对这类白血病的标准治疗疗效极差。将这种新药用于接受标准治疗后效果不佳的患者,疗效非常显著。

### 疗效一般:常见但不明显

然而,多数治疗措施都没有显著的疗效,需要严格的对照试验对其进行评定。有时,一项治疗措施可能在某些情况下明显有效,在其他情况下疗效就不一定显著。例如,磺胺类药物治疗“致命性疾病”产褥热和脑膜炎球菌性脑膜炎疗效显著,但对于其他疾病(相当一部分患者不使用该药也

### 青霉素奇迹

早在 1943 年,我们奉命前往印度南部班加罗尔市的英属总医院进行友好交往。该市科技研究所所长带我四处参观,令我印象深刻。但真正令我震撼的是遇到了两位年轻的生物学家,他们骄傲地展示着一瓶类似粥状物的液体,声称这就是青霉素。他们说:“我们已经阅读过所有文献”,旁边的培养皿里生长了大量的链球菌,他们保证另一个干净的培养皿中链球菌已被青霉素完全清除了。在当时,我们通过从国内和美国定期送来的杂志对青霉素有所耳闻,但知之甚少,对这种药物充满了期待。

不久后我们收治了一位生命垂危的军校学员,他患有感染性海面窦血栓形成和败血症(一种播散性细菌感染,会累及颅内血管)。好几个外科医生都不愿意接诊。我们尝试了所有的办法,但没有任何效果。

这看似一个疯狂的主意,但还有其他办法吗?医生和护士反复讨论;但都没有答案。而此时,患者的昏迷程度逐渐加重。我们最终决定试一试。

我立即骑车去了研究所。我把情况解释后,他们很快同意给我一瓶稀薄的粥状物。我们用了医院最粗的注射针头,把它装在一个大针筒里,我设法通过肌肉注射把这种东西注入患者体内。第二天清晨患者就主动要求喝茶,他恢复得很顺利。似乎那一针就救了他的命。

*Morris JN. Recalling the miracle that was penicillin:  
two memorable patients.*

*Journal of the Royal Society of Medicine.  
2004; 97: 189-90.*

能痊愈),疗效就很一般了。要评价磺胺类药物治疗这类疾病的疗效,就有必要进行严格的对照试验。20 世纪 30 至 40 年代开展的一些临床试验显示,尽管磺胺类药物在丹毒(一种严重的细菌性皮肤感染)和肺炎的治疗中有一定的价值,但对于猩红热似乎不起作用。

同样,尽管维生素 B<sub>12</sub> 治疗恶性贫血疗效确切,但对于患者是否需要每月或每季度注射,至今仍有争论。只有通过比较这两种治疗措施的严格对照试验才能回答这一问题。再者,髌关节置换术后的镇痛效果是显著的,不同类型人工髌关节之间的相对优势十分微妙,但就是这些细微的差别可能非常重要。例如,某些类型的人工髌关节更容易引起髌臼磨损。

### 医生意见出现分歧

有时针对同一种疾病,治疗措施却五花八门,这提示对于最佳治疗措施尚存不确定性。再举一个例子,良性前列腺增生(BPH)的治疗。BPH 是一种好发于老年男性的常见病。最令人烦恼的症状是尿频和排尿困难。治疗 BPH 有几种方案,包括让其自然发展,有时称作“随访观察”或“积极监测”,

这是因为 BPH 症状可能自行缓解。其他非手术治疗包括传统药物治疗(目前 BPH 的用药很多)和天然植物药锯叶棕桐(*Serenoa repens*),这是美国沙苑子或矮棕榈的一种提取物。此外,也可采用手术治疗,不同地区 BPH 的手术率差异很大。例如,在美国,1996 年研究者就国内各地区的 BPH 手术率搜集证据,他们发现各地区的手术率相差近四倍:在美国最大的医疗保险计划 Medicare 的参保人群中,从每千人 6 例到 23 例不等。

我们知道 BPH 治疗措施的选择取决于患者的意愿和利弊权衡:不同患者在评估治疗的风险和效益时可能持有不同的观点。对于那些选择手术的患者,症状缓解的几率最大,但会面临出现并发症的风险,如小便失禁、逆行射精和阳痿。药物治疗在缓解症状上可能欠佳,但却避免了手术并发症的风险。随访观察避免了与手术和药物治疗相关的风险,但很可能无法改善症状。这就需要开展对照试验来比较这些方案,从而帮助患者制定治疗决策。当患者前来就诊时,医生需要解释每个治疗方案的利与弊,共同决定最适合这位患者的治疗方案。可以假设美国各地区人口的男性比例大致相

### 前列腺困境

先生:

对于前列腺癌的筛查、诊断和治疗,有一点是可以肯定的:其筛查、诊断和治疗都尚存不确定性。前列腺癌对于某些患者可能危及生命,而对于其他许多患者并非如此。因此,虽然它可能“可以治愈”,但可能根本就无需治愈。

近来的信函反映,患者在基于证据与专家意见做出治疗决策时有多么艰难。每周都有数百名患者面临着这样的选择。患者 Jeremy Laurance 说他目前暂时不会接受前列腺手术,虽然他的确看起来是在通过忽略这些症状来否定癌症的诊断,其实他只需要去一趟全科医生那就可能明确是良性的可治疗的前列腺疾病而非前列腺癌。Robert Eisenthal 教授却很庆幸自己切除了前列腺。除了说 Jeremy 可能对全科医生的帮助不感兴趣外,没有多少泌尿科医生或肿瘤科医生会指明谁对谁错。在确凿的科学证据出现之前,每周面临这些困境的患者和医生们能做的也只有这些了。

关键在于患者要知情或尽可能知情。减少不确定性的唯一方法是对相关研究给予长期支持。同时,我们需要进一步完善有关男性健康的政策(目前这还不是政府的一个具体目标),帮助他们参与自己的健康管理,用 Wanless 一篇关于公众健康的论文中的话来说就是:要有“健康素养”。

*Dr Chris Hiley, Head of Policy and Research, The  
Prostate Cancer Charity Hiley C.*

*Prostate dilemma [Letter]. The Independent,  
2004, Jun 7, p26.*

同,对于不同个体患者的最佳治疗方案可能会有所不同,但手术率的巨大差异说明医生仍不确定是否应该进行手术、何时进行手术。

如第2章所述,对于是否应该采用前列腺癌的筛查试验也存在相当大的不确定性,从而导致治疗方案的选择上也存在不确定性。

专业上的不确定性还出现在扁桃体摘除术治疗慢性(持续性)扁桃体炎或反复发作的急性扁桃体炎。在过去,儿童几乎是常规施行扁桃体摘除,很少考虑其症状是否值得做手术。如今,扁桃体摘除术的施行更具选择性,但仍常见于儿童,成人中用得也越来越多。然而,目前对于手术适应症尚无统一意见。扁桃体摘除术适用于那些因扁桃体增大而严重影响到呼吸的儿童。但在很多国家,众多的复发性急性扁桃体炎、慢性扁桃体炎及复发性非特异性咽喉炎患者也摘除了扁桃体。尽管“感染”通常被视为施行手术的依据,但不同患者的感染频率和严重程度大不相同。此外,手术无疑是存在风险的,至少全身麻醉就会有风险,包括严重出血。鉴于扁桃体摘除术使用范围上的明显差异,一组研究者决定对已有的对照试验进行系统评价来消除这种不确定性。他们发现尚未针对成人扁桃体摘除术开展任何临床试验。针对儿童的临床试验也不太好:研究者只找到了两项研究,并且都有一定的局限性。例如在其中一项研究中,手术组儿童(仅施行扁桃体摘除术或扁桃体摘除术加增殖腺摘除术)与非手术组儿童存在差异:手术组儿童先前咽喉感染的类型不一样,并且多来自于贫穷家庭。显然,这并非同类比较。研究者因此得出结论:扁桃体摘除术的效果尚未被充分评估,仍需开展更多的对照试验。

乳腺癌的治疗(第2章)又是一个有关医疗不确定性的典型例子。尽管多年来已开展了许多有关乳腺癌诊断和治疗的研究,但在乳腺筛查X片的判读、手术、放疗和化疗的使用方面仍存在巨大差异,提示还有诸多不确定性。对于乳腺癌的基础生物学特征,如基因和酶的功能、患者间的代谢差异,仍有大量尚未解答的问题。极早期乳腺癌患者及癌前患者的最佳治疗方案尚未确定,腋窝淋巴结理想的摘除个数也未明确。对于乳腺癌筛查和治疗的方案尚存争议,因此需要更多的证据来指导实践。除此之外,还有诸多问题尚

### 地域决定命运

“对于医疗保健,常常是地域决定命运”。例如,在美国佛蒙特州,一个社区中8%的儿童摘除了扁桃体,而另一个社区则高达70%。在缅因州,不同社区中70岁以下妇女接受过子宫切除术的比例差异很大,从不足20%到大于70%之间不等。在衣阿华州,85岁以下男性接受过前列腺手术的比例在15%到大于60%之间不等。

Gigerenzer G. *Reckoning with risk: learning to live with uncertainty.*

London: Penguin Books, 2002, p101.

### 医生谈论处方中的猜测成分

两个医生无意中谈到:“我们做的许多事情都是猜的,我相信您我都不好受。要证明治疗是否有效,唯一的办法就是开展恰当的试验,但这样工程就浩大了。那我们该怎么办呢?我们做我们想做的。我肯定有时候这样做是没问题的,这依靠于我们的临床经验、医疗条件等等。然而有时候我们很可能会把错的当成对的,可是由于我们做的事情不能称为试验,因此没人来规范,我们也不能从中学到什么。”

摘自 *Petit-Zeman S. Doctor, what's wrong? Making the NHS human again.*

London: Routledge, 2005, pp79-80.

未得以充分验证,如患者特别关注的结局(如是否能够缓解治疗带来的乏力)或淋巴水肿(手术或放疗后致腋窝疼痛和失功)的最佳治疗方案。

那么,我们该怎么做?首先,医生需要掌握针对某项治疗措施的当前最佳证据,这些证据来自于长期积累的经验以及对现有可靠研究的系统评价。然后医生应与患者讨论各种备选治疗方案,同时尽可能了解患者的意愿。如果这样做还是存在不确定性,医生就应承认不确定性的确存在,并准备好向患者解释为什么会这样。为了使治疗措施更加恰当、更加安全,我们应将这些不确定性视为前进的动力,而不是简单地承认“失败”。患者和医生必须共同努力,设计出更好的研究(第7章)。同时患者需要明白,如果在了解了相关证据后,医生说“我不知道”,您不能因此就去找另外一位医生获得“我知道”的肯定答案,这位医生明显就忽略了实际存在的不确定性。

我们应怎样处理关于某种新药或新技术的这些重要的不确定性呢?显然应该尽力减少不确定性,若要采用一种新的医护方式,必须已有针对该医护方式的研究对其效果进行了评价。一位医学伦理学家这样说道:

“如果我们对某些治疗措施的实际相对疗效不确定,那么我们在将其用于某一适应症时就不能肯定其疗效,也不能随意用于治疗个体患者。因此,在相应的严格试验完成之前,坚持使用某种治疗方式不理智也不合伦理的。由此看来,对于问题“该患者的最佳治疗方案是什么?”,答案就是“试验”。试验即治疗方案。这是一种尝试吗?是的。但我们的意思是在不确定的情况下作出选择,并搜集数据。作出的选择是否“随机”,这点很重要吗?从逻辑上说,不重要。毕竟,在不确定的情况下什么是更好作出选择的机制?”

如果连这样的试验都没有,至少应该以标准的方式记录下使用新的未经验证的治疗措施后出现的结果,这样有助于了解有关该治疗措施的信息,这有益于那些接受了未经验证的治疗措施的患者以及其他所有患者。既然政府将数十亿英镑的税收投入了NHS信息系统,当然有理由希望这些资金是以这种方式服务于公众利益的。

### 患者如何看待不确定性

2002年,美国新泽西州的一名58岁妇女说道:

“可悲的是,从我的个人经历来看我从来没有像《纽约时报》上所说的那样,一个患者应与医生有理性的对话,自己对于乳房X线摄片的担忧应该得到重视。当我对医生说我对每年做一次乳房X线摄片持保留意见时,医生立马就对我充满敌意。结果是我发现我和医生的关系不好,而这对我是不利的。一位优秀的科学家不应惧怕公开解释不确定的问题或讨论问题。恐怕我遇到的那个医生不具备这种科学的思维。”

Diane Palacios 写给美国德克萨斯大学 Anderson 癌症中心 D.A.Berry 医生的信函,2002(经许可后转载)。

Palacios 女士也直接写信给了《纽约时报》,她强调:“我能接受不确定性,但不能忍受不诚实。”

*Palacios D. Re: Senators hear from experts, then support mammography (news article March 1). New York Times, 2002 Mar 4, pA20*

2005年,英国一名低年资医生在自己生病成为患者后写道:

“在NHS(英国国立卫生服务部)的工作经历让我真实地了解了等待、诊断的不确定性以及医生背负的时间压力。我最看重的是对我施行诊断和治疗的医生必须是我所信赖的,也就是说医生应和盘托出已知的相关事实和不确定性,并告知我下一步的计划,而不会只拣好听的说,或者用谈论可能出差错的事来分散我的注意力,或是不停地问我的感受。……我对我父亲(一名儿科医生)说:‘我是幸运的’;但他说我不是,他认为医生本来就应该这样做。”

*Chambers C. Book review. Hippocratic oaths – medicine and its discontents. Journal of the Royal Society of Medicine. 2005; 98: 39-40.*

也有很多医务人员采用了这种负责任的方法。20世纪80年代,英国和加拿大产科医生认为除非了解了绒毛膜绒毛采样(一种用于诊断出生前胎儿畸形的新型有创技术)相对于其他备选方法(如羊膜腔穿刺术)的安全性,他们才会在对照试验中运用这种技术(第7章)。同样,20世纪90年代,英国儿科医生认为只能在随机对照试验中,将新型心肺复苏机用于出生后即缺氧的婴儿。这样医生和患者就会很快发现,在降低婴儿死亡率或严重残疾率方面,新型机器是否优于目前最佳的标准治疗方案。近来,儿科医生已通过相同的方法评价了降低新生患儿的体温将有助于避免脑损伤的建议是否正确。医生采用了这种正规的方法说明他们已经

识到,当治疗措施的效果存在不确定性时,即便是出于一片好心,也很容易给患者带来意外的伤害。

尽管现在越来越多的人都支持这种做法,特别是对于那些新的且通常较昂贵的治疗措施,然而遗憾的是,我们还是错过了很多机会。例如,目前尚无针对可迅速致命的毁灭性神经系统疾病克雅氏病(其中一种类型很可能是因食用了疯牛病牛肉所致)的有效治疗方法。可以理解的是,患者家属有时会要求试用任何有希望的治疗,即便对该治疗措施的利弊平衡关系所知甚少,甚至一无所知(第7章)。但是如果这些治疗措施已经过了公平试验的充分评价,患者将会受益更多。

### 减少治疗措施效果的不确定性

要尽快有效减少关于治疗措施效果的不确定性,就必须采取一些措施。我们将在本书最后两章讨论部分措施,尤其是患者的参与。但在此我们还想特别指出一点,当关于治疗措施效果的信息不足时,应确保医生在获知该治疗措施的价值和潜在危害之前,只能在正规评价中使用这一措施,以此来加深对该治疗措施的了解。但为什么有人会反对这种可以降低风险的方法呢?

### 谁说医学研究对您的健康有害?

“大多有关医学研究的伦理讨论都涉及了这样一个问题:应如何规范医学研究?事实上,从很多方面来看,医学研究相对于医学实践有更加严格的规范。细读有关医学研究的大量指南后,您就会原谅医学研究可能对您的健康产生的危害,这就和吸烟的感受是一样的。”

*Hope T. Medical ethics: a very short introduce. Oxford: Oxford University Press, p99.*

30年前,这一问题就困扰了一名英国儿科医生,他发现要给他一半的患者施行某种治疗就需要得到许可(给一半的患者施行已有的治疗措施,并以此作为对照,评价给另一半患者施行的新治疗措施的疗效),但如果他要把该治疗作为标准治疗方案用于所有患者,却无需任何许可。这种不合逻辑的双重标准至今仍然存在,也打消了部分临床医生希望减少治疗效果不确定性的念头。例如,英国医疗委员会(GMC)建议医生:“您给每位患者提供的信息量应该有所不同,这取决于多种因素,如疾病的性质、治疗方案的复杂程度、与治疗或手术相关的风险、患者自己的意愿等。”但GMC关于医学研究知情同意的指南却是这样要求的:“受邀参与临床试验的患者必须全面获知所有的相关信息。”

考虑每一位接受了治疗的患者的利益,而不能仅考虑少数临床试验参与者的利益,这点很重要也合乎伦理。显然,所有相关信息都应按需提供,但也可以根据个体患者的意愿和要求(可能随时间而发生变化)提供部分信息。固执地坚持提供所有信息或在任何情况下都要获取知情同意可能有悖于常识和规范的临床实践。

无论是否在参与临床试验,患者都需要治疗与关怀,由于患者的需求、理解能力、焦急或恐惧程度的不同,他们得到的信息量可能也会有所不同。当医生试图通过告知患者所有相关信息来减少有关治疗疗效的不确定性时,这往往会使患者感到非常不安,他们更愿意“把这些烦恼丢给医生”。加强医务人员的沟通能力似乎比执意坚持提供所有信息更有效。医患之间的相互信任才是有效沟通的基础,因此灵活的处理方式可能更加有用。GMC目前正在修订其知情同意指南,将建议每位参与临床试验的受试者都有机会轻松获取所有相关信息。

### 有偏倚的伦理学

只要临床医生尝试了某种新的疗法,希望对其进行仔细研究、评价结局并发表结果。这就是在做研究。这类研究的受试者就应该得到特殊保护。研究方案必须通过机构审查委员会( IRB ) [ 在英国相当于伦理委员会 ] 的批准。知情同意书必须经过严格的审查,部分研究可能会因此被禁止。但另一方面,临床医生可能根本没有出于研究的心态而尝试了新的疗法,仅仅是因为他相信新疗法对患者有用。在这种情况下,这种尝试就不算是研究,也不需要 IRB 批准,可能只是出于避免医疗纠纷而获取知情同意。

不难看出,患者在第二种情况下(非研究)承担的风险远远大于第一种情况(有正式医学研究的成分)。此外,在第一种情况下,医生更具有伦理学观念。在第一种情况下医生是在对治疗进行评价,而在第二种情况下医生是在根据他并不可靠的预感试用治疗。然而,由于旨在保护患者的伦理学规范重在创造更多有益于公众健康的知识,但这只能规范那些有责任的研究者,而无法约束到那些不负责任的冒险者。

Lantos J. Ethical issues-how can we distinguish clinical research from innovative therapy?  
American Journal of Pediatric Hematology/Oncology.  
1994; 16: 72-5.

要清楚地解释治疗效果的不确定性,从医生方面来讲需要一定的技巧和谦逊态度。当许多医生对可能参与试验的患者解释目前尚不知道哪种治疗最好时,都觉得很难。但是,现在公众的态度改变了:那些以“上帝”自居的傲慢医生越来越不被患者所喜爱。我们应该着重训练那些不羞于承认自己也是凡人,也需要帮助的医生,需要患者参与研究以进一步确定治疗措施的疗效。

许多医务人员和患者遇到的主要问题就是不理解什么是随机、为什么需要随机、随机的通俗含义是什么(第3章)。这提示急需提供医生和患者都能获取的清楚的临床试验书面信息,并阐明为什么需要这么做。

这涉及两方面的不同需求:第一,通过普及教育使公众

### 互动的个性化知情同意方法

“早上好, Jones 女士,我是 Smith 教授。请坐。您的医生可能已经向您提过他希望我来看看您,因为您气喘的问题一直没有好转,看我是否能有一些办法来帮助您。我希望我能帮助您,但这可能需要您在未来几个月内常过来就诊,我们共同探讨最适合您的治疗方案。”

“如果我能更多地了解您,包括您的病情、您的想法和意愿,我就可以更多地帮助您。由于这是我们第一次见面,我想先简单说一下我的想法。不同的患者愿意提供给医生的信息量以及从医生那里获取的信息量都是不同的。大部分患者都觉得自己从医生那里得到的信息不够;但有的患者却宁愿自己少知道一些。由于我们现在都还不了解对方,所以我希望您告诉我,对于您的疾病以及可选的治疗方法您想了解多少。如果您认为从我这里得到的信息过少或过多了,您可以随时提醒我。还有一点需要您明确,在回答您提出的任何直接问题时,我绝不会撒谎。如果我不知道答案,我会尽量为您找到答案。您认为这种方式您可以接受吗?”

Oxman AD, Chalmers I, Sackett DL. A practical guide to informed consent to treatment.  
British Medical Journal. 2001; 323: 1464-6.

理解什么是随机对照试验以及为什么要做这些试验;第二,解释在特定的临床试验中为什么要给患者提供这种治疗。在医院的候诊区内应摆放为什么需要临床试验的宣传单。这有助于患者理解临床试验,避免患者在受邀参加临床试验时感到很惊讶。总之,临床试验在医务人员和患者之间应该成为一种伙伴合作关系,这样才有助于确定治疗方案、提高生活质量。然后,请患者参加临床试验告诉他们试验是怎么做的以及为什么要邀请他们参与。我们非常乐意看到有那么一天:患者在受邀参加临床试验时一点也不感到奇怪,他们还会主动询问自己可以参加哪些试验,并且能够判断所研究的问题对自己是否重要。

### 随机——一个简单的解释

“随机就是为了最大限度地减小偏倚并确保各组患者所有已知的和未知的影响因素上尽可能相似。这样才能确保结局指标的组间差异是因治疗措施疗效的差异所致,而不是因患者的差异所致。”

“这避免了医生有意或无意地将某种治疗分配给某类特殊患者,而将另一种治疗分配给另一类患者,同时也避免了某一类患者选用某种治疗,而另一类患者选用另一种治疗。”

Harrison J. Presentation to Consumers' Advisory Group for Clinical Trials, 1995.

## 伦理学、医学研究伦理委员会与患者利益

奇怪的是,当治疗措施的效果存在不确定性时,医学伦理学家和医学研究伦理委员会竟然都支持这个双重标准。伦理学家通常更注重保护“弱势群体”,较少鼓励患者也合作参与临床试验。但正如一位医学伦理学家所说:“如果伦理学家或其他人要给临床试验挑错的话,他们会去关注哪些工作不够科学、是否在做重复研究,受试者的排除是否不当、资源的使用是否不公平或不合理。这其实都忽略了一个问题,试验的根本目的是什么,是为了证明我们使用的治疗措施是安全的,并且比其他治疗更有效。伦理学上没有捷径,临床试验也是一样。”

自19世纪30年代起,由于人体试验多次出现了明显不道德的各种丑闻,从而诞生了医学研究伦理委员会(评价新开展的研究的伦理问题的独立委员会)。这类委员会对于保护患者不会以研究的名义而受到伤害以及严格审查研究的设计是否能够提升科学认识等方面非常重要。这类委员会的主要目的并不是评价治疗措施的疗效。因此,就临床试验提供给患者的治疗措施而言,伦理委员会能够发挥的作用就很小了。

### 以患者为导向的对照试验指南

目前,国际对照试验 meta 分析注册中心已经建立 ([www.controlled-trials.com](http://www.controlled-trials.com), accessed, Aug 9, 2000), 其基础工作之一是制定一份以用户为导向的对照试验指南(电子版),以帮助考虑参加临床试验的患者在充分知情的情况下做出选择。网站上公布了用户对注册试验的评价,例如所研究的问题是否重要、以前的研究是否已经充分解答了这个问题、研究的设计是否科学、是否合乎伦理、所选用的主要结局指标对患者是否重要、受试者是否能够清楚获知研究结果等。用户的意见将有助于改进临床研究方案,更好地服务于患者,正如 Sheila Kitzinger 制定的新生儿指南,有力地促进了英国产科医院更加注意各自的公众形象。

*Chalmers L. A patient-led Good Controlled Trials Guide, Lancet 2000; 356: 774.*

- 伦理委员会尚不能有效区分医学研究评价的是几乎未用过的治疗措施(事实上部分治疗药物可能还没有上市)的效果还是那些已经广泛使用的治疗措施的效果。

- 上文提到的知情同意双重标准的问题,伦理委员会还不够重视,甚至根本没有重视。

- 伦理委员会没有要求在提出新的研究方案之前对已有研究进行系统评价。

- 伦理委员会没有要求研究者上报研究的利益冲突。

- 伦理委员会在避免研究结果报告不足(报告偏倚)的问题上没有起到任何作用。

因此,伦理委员会在为公众利益服务方面存在严重的问题。NHS 采取了一些措施确保已批准的研究得到的结果必须全部公之于众。例如,每个伦理委员会应该对其审查的每一项试验登记,并要求研究者在试验结束后三个月内提交最终的研究报告。但仍然还有许多问题需要解决,以确保伦理委员会的工作充分且透明。只有这样,患者在受邀参加治疗性临床试验时才能够确信该研究是有价值的,他们所做出的努力和牺牲是有用的。

### 临床研究与实践密不可分

从伦理学的角度来看,临床研究和临床实践应该是一致的。无论对于旨在帮助患者的非常新的治疗方法(未知其潜在的利和弊),还是对于已经认可的治疗方法(可能经验多一些,但其价值尚未被证实)都如此。对于 NHS,从伦理学上应该鼓励开展科学研究以发现如何让有限的资源发挥最大的效益。

*Advisory Group on Health Technology Assessment.  
Assessing the effects of Health Technologies:  
Principles, Practice, Proposals. London: Department  
of Health, 1992, p25.*

### 要点

- 治疗措施效果的不确定性是普遍存在的;
- 当我们发现没有人知道有关疗效的不确定性时,应该采取措施来减少这种不确定性;
- 有必要认真记录和评价以及严格实施临床试验;
- 临床试验和非临床试验的知情同意双重标准不符合患者的利益;
- 必须进一步采取措施让患者共同参与减少治疗措施效果的不确定性;
- 医学研究伦理委员会目前并不能有助于患者。

## 第5章 临床研究：高质量研究，低质量研究和不必要的研究

前几章我们强调了为什么研究需要严格的设计并且必需解决关乎患者和公众的问题。只要做到了这些，即便没有出现预期的结果，我们仍会感到骄傲和满足，因为我们已经取得重大的认识并减少了不确定性。许多临床研究由于遵从了研究设计和报告标准而成为高质量研究。然而，由于各种原因，低质量的和不必要的研究仍在开展和发表。

### 高质量研究

中风是导致患者死亡和长期残疾的主要原因。首次中风后的死亡率约为 1/6 ~ 2/6，二次中风后的死亡率则上升到 4/6，二次中风多出现在首次中风后一年内并累及同一大脑区域。向大脑供血的颈动脉狭窄是导致中风的潜在原因之一。有时黏附在颈动脉内膜上的脂性物质脱落，栓塞小动脉而引起中风。20 世纪 50 年代，外科医生开始采用颈动脉内膜剥离术来清除这类脂性沉积物，目的是希望手术能减少出现中风的风险。然而，任何手术操作本身都有引起并发症的风险。

尽管颈动脉内膜剥离术越来越风行，但直到 20 世纪 80 年代才进行了随机对照试验来评价手术的利弊。显然，该信息对患者和医生都极其重要。两项设计精良的试验（一项在欧洲，另一项在北美）比较了手术治疗与当时最有效的非手术治疗，受试人群是已经出现了颈动脉狭窄症状的患者（小中风或短暂的中风样症状）。数千名患者参与了这两项长期试验。于 20 世纪 90 年代发表的试验结果显示，手术可降低中风或死亡风险，但这取决于颈动脉狭窄的程度。狭窄程度相对较轻的患者接受手术治疗则弊大于利，因为手术本身也可能引起中风。这些重要的发现直接影响了临床实践。

另一个高质量研究的突出范例是关于妊娠并发症的。全球每年约有 60 万妇女死于妊娠并发症。死亡病例多出现在发展中国家，并且多数与妊娠惊厥或抽搐（称为子痫）有关。子痫是一种致命性疾病，可导致母婴死亡。妇女出现先兆子痫（也称为毒血症）的前兆症状是高血压和蛋白尿。

1995 年，研究显示注射硫酸镁（一种价格低廉的常用药）可以预防患有子痫的妇女再次出现抽搐（参见第 6 章）。研究还显示硫酸镁控制惊厥的效果优于其它标准抗惊厥药物。因此研究者意识到确定硫酸镁是否能预防患有先兆子痫的妇女出现惊厥非常重要。为此开展了 Magpie 试验，来自全球 33 个国家的万余名患有先兆子痫的妊娠期妇女参与了该试验。除了常规医疗护理外，一半的受试者接受硫酸镁注射，另一半则接受安慰剂（模拟药物）。Magpie 试验得出了明确且令人信服的结果，即硫酸镁使出现惊厥的风险至少降低了一半。另外，虽然硫酸镁没有明显降低婴儿的死亡风险，但有证据提示其能够降低母亲的死亡风险。除外轻微的副作用，硫酸镁对母婴是无害的。

### 我参与 Magpie 试验的经历

“我很高兴能够参与一项如此重要的试验。我在孕 32 周时出现了水肿，并且越来越严重，后来被诊断为患有先兆子痫，在孕 38 周时开始入院治疗。我接受的是剖腹产，谢天谢地我和孩子都完全康复了。先兆子痫是一种可怕的疾病，我由衷希望该试验的结果能惠及象我这样的患者。”

Clair Giles, Magpie 试验的参与者

MRC News Release. Magnesium sulphate halves risk of eclampsia and can save lives of pregnant women. London: MRC,

May 31, 2002

高质量研究的结果还改变了人免疫缺陷病毒（HIV，AIDS 的病因）感染儿童的命运。全世界每天约有千余名儿童死于 HIV 感染或与 AIDS 相关的疾病。细菌感染是常见的死因，比如肺炎，这与儿童免疫系统功能低下有关。复方新诺明是一种价格低廉的常见抗菌药，多年来一直用于治疗与 AIDS 无关的儿童或成人胸部感染。在感染了 HIV 的成人中开展的研究表明，该药可减少细菌感染引起的其它并发症。

最初的证据显示，复方新诺明可降低携带 HIV 的儿童感染率，英国与赞比亚的研究者合作，在赞比亚开展了一项大型试验以检测复方新诺明作为预防药物的效果。该试验于 2001 年启动，持续了约 2 年，在 500 余名儿童中将这种抗菌药与安慰剂进行了比较。很快就得到了明确的结果，即该药使与 AIDS 相关的死亡减少了 43%，需要住院治疗的人数减少了 23%。此时，独立委员会在审查了试验结果后建议终止试验。所产生的直接结果是赞比亚政府要求所有参与该试验的儿童均接受复方新诺明治疗。世界卫生组织和联合国儿童基金会立即改变了其关于携带 HIV 的儿童有效治疗药物的推荐意见。

### 低质量研究

遗憾的是，研究的质量并非都很高。以迟发性运动障碍为例，这是一种令人烦恼的疾病，是由长期服用神经镇定药（抗精神病药物）导致的严重副作用。神经镇定药用于治疗精神疾病，特别是精神分裂症。迟发性运动障碍最突出的特征是反复出现不自主运动，如作鬼脸、舔唇、频繁吐舌、皱额或鼓腮，有时还伴有手脚抽搐。服用一种神经镇定药超过 3 个月，五分之一的患者都会出现这种副作用。

20 世纪 90 年代，一组研究者开始系统评价过去 30 年内迟发性运动障碍的治疗措施。让他们感到惊讶的是共找到了约 500 项随机对照试验，共涉及 90 种不同的药物治疗（检索截至 1996 年）。但这些试验都没有得出有用的数据。部分试验因样本量太小导致结果不可靠；另一些试验则因疗程太短导致治疗无意义。

该研究小组还发表了有关精神分裂症治疗措施的对照试验的内容和质量的综合评价结果。研究者共评价了 2000 项试验,结果令他们感到失望。从某些方面来看,过去几年内药物的确改善了精神分裂症患者的预后与转归。例如多数患者现在能在家或社区生活。然而,即使在 20 世纪 90 年代甚至在今天,多数药物试验却是在住院患者中进行的,因此无法确定其与门诊治疗的相关性。更严重的是,治疗结局的评价方法也存在惊人的差异。研究者发现这些试验共评价了 600 多种治疗措施(主要是药物治疗,也有心理治疗),却采用了 640 种不同的量表来评价结果,其中 369 种仅使用了一次。因而无法比较不同试验的结果,医生和患者也难以理解这些试验结果的实际意义。其它问题还包括许多试验样本量太小或试验周期太短而无法得到有用的结果。新药往往是与有明显副作用的药物进行比较,这显然是不公平的试验。研究者最终得出结论:半个世纪以来的试验在质量、试验周期和临床相关性等方面都不尽如人意,急需更多设计精良、严格实施且报告充分的试验。

#### 更知情

“即使缺乏有关干预措施有效性的证据,或证据质量较低,也必须避免臆断猜测。通过制作系统评价并指出是否存在高质量证据,至少可以让临床医生、研究者、决策者和患者更知情。”

Soares K, McGrath J, Adams C. Evidence and tardive dyskinesia. *Lancet* 1996; 347: 1696-7.

早年关于硬膜外镇痛法缓解妇女产时疼痛的试验明确提示必须评价与患者相关的结局。20 世纪 90 年代,研究者通过对照试验比较了硬膜外与非硬膜外镇痛法。结果显示,近 20 年来,尽管有数百万的妇女都接受了硬膜外阻滞,但参与了将硬膜外阻滞与其它形式的镇痛方法进行无偏倚对照的试验的妇女还不及 600 名。研究者共检索到了 9 项对照试验并对其进行分析:这些试验大多测量的是激素水平以及可反映产时应激状态的其它物质的水平。部分试验也评价了婴儿的结局。然而,除 2 项试验外,其他所有试验都没有比较产妇自我报告的疼痛程度。换言之,这些试验的执行业者竟然忽略了一个至关重要的结局,即产妇的疼痛是否得以缓解。

第 3 章解释了选择性报告研究结果可导致严重偏倚。一些“阴性”研究结果因没有达到研究者或资助者的期望而被隐藏。这些试验没有公布试验报告就消失得无影无踪了。另一方面,已经发表的试验也可能选择性地报告结果,也就是说部分结果若不支持研究者或资助者对所验证治疗措施疗效的期望,就不会被报告。这种做法是站不住脚的。但为什么有偏倚的报告不全问题如此重要?

2004 年,一组研究者公布了对这类报告偏倚的首次全面评价结果。他们共分析了 100 多项随机试验,并获取了相

应试验方案的副本以及所有方案修订信息,从而可以了解试验计划评价哪些结局。然而对照相应的试验报告结果令他们感到吃惊。他们发现报告不全的问题严重,其中一半涉及治疗措施的有效效应,三分之二则涉及有害效应。

研究者后来又向相关的试验者发送了问卷,质询未报告的结果。但问卷的回收率却不到 50%,其中绝大多数人一开始都否认漏报了试验结果,即便从最初的试验方案中可以找到确凿的证据,有时甚至在已发表文章的方法部分都提到了这些未报告的结局。为了杜绝这种错误的做法,研究者呼吁对所有待开展的试验注册,并且试验方案应向公众公布,以接受监督。

#### 低质量研究的影响

“除了选择性发表外,结局的报告偏倚也产生着深远的影响。它使得不真实的研究结果泛滥,从而导致文献综述过高估计干预措施的效果。若将无效的甚至是有害的干预措施大肆宣传,这对患者、医务人员和决策者就再糟不过了。另外还有一个问题,将昂贵的治疗措施大力推广,声称其比价廉的治疗措施效果更好,然而事实却并非如此。”

Chan A-W, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *Journal of the American Medical Association* 2005;2457-65

#### 不必要的研究

还有一些研究介于高质量与低质量研究之间,即显然是不必要的研究。比如关于早产儿的研究,早产的婴儿肺发育不全,可能出现危及生命的并发症,如呼吸窘迫综合征。早在 20 世纪 80 年代初期,就有证据明确提示存在早产风险的孕妇使用类固醇药物可降低新生儿呼吸窘迫综合征的发生率和死亡率。然而此后 10 年内,仍在继续开展比较类固醇与安慰剂或空白对照的试验。如果对早期试验进行系统评价并通过 Meta 分析(第 3 章)合并结果,那么就不会再开展后来的那些试验了,因为早期试验的综合证据即提示根本没有必要再做试验。

第 1 章也提到过另一个不必要研究的例子,也是因为没有收集已有研究结果并对其进行分析所致。该临床试验是在中风患者中验证药物尼莫地平是否能减轻脑损伤的程度。但之前却从来没有对相关动物实验的结果进行过系统评价和严格分析。如果先进行了评价和分析,就会发现动物实验存在诸多显而易见的问题,如没有随机化、没有采用盲法评价结局等。以至后来研究者却基于这些所谓“鼓舞人心”的动物实验结果在中风患者中开展了临床试验。

另一个例子是关于肠道手术的。1969 年,一项试验将抗生素与安慰剂(模拟药物)进行比较,评价其是否能降低

术后死亡率。这是一项小样本试验,没有得出肯定的结论。因此 20 世纪 70 年代又开展了进一步的研究来减少这种不确定性。随着证据的增多,到 70 年代中期就已明确抗生素的确能够减少术后死亡。即便如此,80 年代伦理委员会仍在继续批准涉及相同问题的试验,研究者也仍在实施这些试验。因此,在这些试验中有一半的患者都没能接受已证实能够减少术后死亡的治疗措施。为什么会这样呢?最可能的原因是研究者没有对之前累积的证据进行系统评价,或没有参照合并了所有相关证据的最新系统评价就启动了新的试

验。显然,伦理委员会没有要求研究者去做这些工作就批准了新的试验。换言之,研究者和伦理委员会都没有把患者的利益放在首位。

#### 要 点

- 不必要的研究等于浪费时间、精力、金钱和其它资源;
- 只有当最新的系统评价提示有必要进一步研究时,才能开展新的研究;
- 新的研究证据应及时用于系统评价的更新。

## 第 6 章 精简、优质且出于正当理由的研究

多年前,《英国医学杂志》(BMJ)刊登的一篇社论标题醒目:“低质量医学研究的丑闻”。作者提倡精简、优质且出于正当理由的研究。前几章已经提到了困扰该作者的一些问题。

### 精简研究

生物医学研究机构以及开展这类研究的研究者大多宣称其研究目标很简单,即为促进人类健康提供有用的信息。但在每年发表的数以百万计的生物医学研究报告中,又有多少真正为这项事业做出了贡献呢?

#### “信息山”

“每年有超过 2 百万篇生物医学文献在 2 万余种杂志上发表——几乎就是一座‘信息山’……把这些文献堆起来可高达 500 米。”

Mulrow CD. *Rational for systematic reviews.*

In: Chalmers I, Altman CD eds. *Systematic reviews.* London:

BMJ Books, 1995.

英国布里斯托的研究者提出了这样一个根本问题:“有关关节炎的研究在多大程度上反映了该病患者及其主管医生所关注的重要问题?”他们召集了 4 组人,分别是患者、风湿病学家、理疗师和全科医生。这 4 组人一致表示不希望再看到那些由药商资助的比较某种非甾体类抗炎药(如布洛芬)与安慰剂(模拟药物)的药物试验。他们更希望看到的是将理疗与外科手术进行严格比较以及对教育和处理策略进行评价的研究,这些研究才可能帮助患者更好地管理这种慢性、致残性且令人痛苦的疾病。当然,这类治疗和管理措施的商业价值比药物治疗要小得多,所以常常被忽略。

若按此方法评估,又有多少其他领域的治疗性研究也存在类似的研究者所做研究与患者和医生所关注的问题不相符的现象?其他事例让我们相信,这并非个案而是普遍现象。药物制剂的微小变化很少引起药物疗效实质性的改善,但这类研究却在关节炎及其他疾病的研究中占据主导地位。这是多么大的资源浪费!

显然,这种现象是不正常的,怎么会这样呢?原因之一是研究者开展研究的初衷被某些外部因素扭曲了。例如,药厂开展研究的首要目的是对其股东负责,而不是要满足患者和医生的需求。药厂的研究大多瞄准大市场,如犹豫着是否要采用激素替代治疗的妇女,或陷入抑郁、焦虑、不安和痛苦中的人们。然而近几十年来,这种以商业为目的的研究,即使瞄准的是“大众市场”,也极少产生重要的新治疗措施。相反,对于同一类药物,药厂常生产出各种非常相似的药物,即“me-too”药(仿制创新药)。这就像当年超市里只供应面包的时候,白切片面包就不断地变着花样。药厂在市场上花

的功夫远比在研究花的功夫要多,这就不足为奇了。

#### ‘me-too’药在加拿大的影响

“在英国哥伦比亚,1996 至 2003 年间大部分(80%)药费增长都是由于使用了新的专利药品,这些新药比 1990 年以前的药品在价格上高很多,但疗效却没有实质性的提高。‘me-too’药的价格远远超过那些经历了时间考验的老药,这笔增加的费用值得我们仔细审视。正如新西兰一样,合理的药物定价可以节省开支,用于满足其他卫生保健需求。例如,如果英国哥伦比亚 2003 年消费的‘me-too’药中有一半的定价与老药相当,就可节省 \$ 350 m (相当于处方药总费用的 26%)。这笔省下的开支可以支付千余名新医生的薪水。

鉴于全球销量居前 20 位的药物包括许多老的药物种类中出现的新专利药……由此看来,在多数发达国家‘me-too’药很可能主导着消费趋势。”

Morgan SG, Bassett KL, Wright JM, et al. “Breakthrough”

drugs and growth in expenditure on prescription drugs in Canada.

British Medical Journal 2005; 331: 815-6.

但是,药厂如何劝说医生放弃便宜的老药改用新药呢?常用的办法是开展许多小型研究项目,证明新药比什么药都不用更好,但这些研究不会去比较新药与老药。遗憾的是,药厂轻而易举就能找到愿意参与这些无意义研究的医生。这些医生最终还会为以这种方式研究出来的药品说好话。药品监管部门一贯坚持新药与安慰剂对照,而不是与已有的有效治疗措施对照,这使得问题更加严重。

诸如《柳叶刀》一类的权威医学杂志曾发表述评警示,当前部分临床研究的参与者动机不纯,医学院校与药厂之间的关系越来越值得怀疑。《新英格兰医学杂志》的一位评论员曾直言不讳:“医学研究难道是可以出售的吗?”

然而,商业利益并不是导致生物医学研究忽视患者利益的唯一负面影响因素。许多医学院校和科研基金组织都认为弄清疾病的发病机理可能更有助于改善健康状况。因此,他们忙于在实验室开展动物实验。虽然不可否认,这些基础研究是必要的,但鲜有证据证明其存在偏倚。实验室研究在开展之前很少仔细评价其对于患者的相关性,最终导致实验室研究过剩。

导致这种扭曲现象的原因之一是研究者都急切地希望取得医学进展,而基础研究最有可能取得进展。例如,发现 DNA 结构后 50 年,人们纷纷开始质疑其给卫生保健带来的潜在效益。正如一位遗传学家所说:“20 年来,遗传学家们对于研究成果做出了太多的承诺,但很少实现,有的永远也不会实现。”基础研究发现治疗措施的治疗理论后,必需在患者中开展设计精良的研究来验证这种治疗理论,除此之外,别无他法。然而,研究者通常都没有去追踪评价这些治疗理论是否的确与患者相关。在研究者发现了导致囊性纤维变

性的基因缺陷后十多年,患有这种病的患者仍在询问,什么时候才能获得该研究成果带给他们的健康利益(第2章)。

### 骗子、傻瓜和坏蛋?

两名研究者为了给《英国医学杂志》(BMJ)的圣诞特刊写一篇幽默诙谐的文章,恶搞了一家名为HARLOT plc的公司,要给试验主办者提供一系列的服务。诸如:

“生产冒牌药品和器械但又想增加其市场份额的厂家们,希望多吃一些不必要的诊断性检查、多提供一些不必要的治疗性服务的医疗机构,以及那些希望推行不理性的、为自我服务的卫生政策的国家和地方卫生部门……还有那些骗人的‘me-too’药……只要你们的‘me-too’药不比三重蒸馏水差很多(我们拥有专业的E-Zee-Me-Too试验方案小组),我们就可以提供试验服务并且保证结果阳性。”

令人惊讶的是,真的有人致信作者认真地询问这个荒谬的HARLOT plc皮包公司的情况。

Sackett DL, Oxman AD. HARLOT plc. An amalgamation of the world's two oldest professions. *British Medical Journal* 2003; 327: 1442-5.

## 优质研究

即便研究看起来似乎与患者相关,但研究者在设计研究时常会忽略患者所关注的问题。就拿一个有力的例证来说,研究者让肺癌专科医生把自己设想成患者,并考虑自己是否会同意参与其自身可能符合纳入标准的6项肺癌试验之一。36%至89%的医生表示不会参与,这大概是因为医生认为研究所探讨的问题不够重要。

同样,银屑病是一种慢性致残性皮肤病,在英国每一百人中就有两人患有此病。但在银屑病的临床试验中,患者的利益并没有得以充分体现。银屑病协会发现,在许多研究中研究者都坚持使用很不可信的评分系统来评估不同治疗措施的效果。其缺点之一为,评分系统强调的是测量结果,例如患处皮肤的总面积和皮损的厚度,然而,患者更关心的却是位于面部、手掌、足底和生殖器的皮损。

那么,自从BMJ的社论揭露了低质量医学研究的丑闻后,这十年来情况有所改善吗?令人高兴的是,答案是肯定的。在研究调查纷纷揭露了许多临床试验报告质量低下的事实后,临床试验的报告标准现已形成并得以推广(第3章)。这类标准要求清楚地报告多项试验相关细节,如邀请了多少患者参与试验,又有多少患者拒绝参与,并根据试验开始时的分组情况报告结果。但在很多方面还做得不够,仍有待改善:(a)研究所要探讨问题的选择;(b)形成这些问题的途径,以确保所评价的治疗结局是患者重点关注的;(c)向患者提供的试验相关信息。

### 研究没能很好地为银屑病患者服务

“鲜有研究将不同的治疗措施进行对照或关注长期治疗方案。对于这种很可能持续终身的慢性疾病,试验周期显然太短。我们唯一能肯定的似乎只是治疗比不治好。显然,研究者完全忽视了患者的经验、观点、意愿和满意度。”

Jobing R. Therapeutic research into psoriasis: patients' perspectives, priorities, and interests. In: Rawins M, Littlejohns P, eds. *Delivering quality in the NHS* 2005. Abingdon: Radcliffe Publishing Ltd, pp53-56.

一些研究者开始逐渐意识到与患者、护理者以及公众合作,共同了解其疾病和卫生保健经历的重要性。社会学家越来越多地参与到验证治疗措施的研究小组中。因此,逐渐形成了探索疾病这些方面的正规方法,并开始引入将传统方法与证据相结合的新途径。当患者受邀参与临床试验时,如果试验所要探讨的问题是患者与研究者共同确定的,患者将更有可能意识到开展研究以减少治疗不确定性是必要的。为了探讨一项试验的可行性和可接受性,可以先开展由多组患者参与的预试验。这种预试验可能会暴露试验设计方案的缺陷;或有助于确定更相关的结局指标,甚至可能提示试验理论一开始就是错误的。这样就节省了大量的时间和金钱,并避免了失败。计划在局限性前列腺癌患者中开展的临床试验的前期准备工作(第7章)给我们上了生动的一课。预试验提示应认真斟酌临床医生描述试验目的和治疗选择所用的术语,以进一步完善试验设计。当癌细胞还没有扩散到前列腺以外时,不能确定的是应该单纯定期监测患者还是建议行手术或放疗。研究显示,许多因心脏病、中风或其它原因死亡的老年男性,其实都还存在局限性前列腺癌,但该病从未影响到他们的健康状况。然而,该临床试验的前期准备工作显示,临床医生在确定治疗这种无症状性疾病的价值上是确有困难的。同样,他们也在如何公平地描述各种治疗选择上犹豫不定。

临床医生也会不自觉地使用可能让患者误解的术语。例如,医生在描述试验的监测组时常会使用“观察等待”这样的术语,患者可能会理解为“不治疗”,似乎医生“到我死的时候才会观察”。因此,研究者将“观察等待”替换为“积极监测”,并详细描述定期监测具体要做什么。此外,由于担心这一组的干预措施不会被患者所接受,医生常拖到最后一刻才跟患者讲明。而且,医生本来想安慰患者有很大机会存活十年以上,但使用一些术语却无意中造成了患者的不安。医生说:“大多数前列腺癌患者可以存活十年以上。”患者却会消极地理解为“部分患者会在十年内死亡”,而非“大部分前列腺癌患者可以带病存活很久。”

因此,该试验认真解读了患者在预试验中的看法,而没有强加试验者的设计观点,从而形成了一种可行且能被患者

所接受的方法来比较各种治疗措施。一项能被患者所接受的试验更有可能在短时间内完成招募工作,从而更快地发现治疗措施的优劣,这对患者和医护人员都是有益的。

### 出于正当理由的研究

由于大量生物医学研究经费是投入到了实验室和动物研究中,因此对于可能快速生产出与患者相关的证据的研究却遭遇着经费危机。因此,药厂在确定应研究有关治疗措施效果的哪些问题时,是将经济利益放在首位的。学者及其所在单位参与试验时通常会按照药厂的要求来行事,因为每纳入一名患者,药厂就会支付他们数千英镑的报酬。有时这些所谓的“赞助费”会用于填补单位的经费;但还有不少流入了医生个人的腰包。

大学里的一些其他奖励制度也导致了出于不正当理由的研究。正如《新英格兰医学杂志》的一位前任编辑所说:“相对于个人或小组研究,大规模多中心临床试验让研究者成为作者的机会更少”。在大学里,成为研究报告的作者会得到高度奖励,这是个人职称晋升的条件之一,也是单位是否成功以及是否能继续得到资助的衡量标准。因此,研究者及其所在单位认为多中心合作研究项目(通常以团体的名义发表)是对他们个人能力的认可和荣誉的一种威胁。所以,研究继续以个人或小组研究的形式进行,从而导致样本量较小,不足以得出对患者很重要的可靠结果。

在英国,政府的大学拨款制度实际上也助长了这一趋势,服务于学术利益而不是患者利益。一位竭力帮助中风患者的资深神经科医生难过地指出:“合作研究项目的署名通常都无法真实反映随机对照试验从构思、设计到施行的整个过程,而署名是对众多试验参与者的一种肯定。学校的一名领导曾斥责我被引用的次数不够,然而作为一名基础科学研究者,同时也作为一名商人,他并不清楚我在试验中所做的工作。”

这种被扭曲的研究动机不得不令人担忧,这不仅关系到已进行的和已发表的研究,还关系到那些根本没有进行的研究。这些趋势的“机会成本”是真实存在的,许多旨在增进健康的关于治疗措施效果的问题尚属研究空白,因为这些问题不能满足药厂或大学的利益。

例如,子痫是一种危及生命的妊娠并发症(第5章),每年导致约5万名妇女死亡,但有关如何控制有惊厥史的妇女再次抽搐的争论竟然持续了将近一个世纪。这一问题没有引起药厂的关注,原因有两个:(a)死于该病的妇女大多生活在发展中国家;(b)治疗药物之一硫酸镁是一种使用历史悠久的泻盐,商业前景不大。这一问题也没有引起学术界的关注,因为只有通过大规模的国际合作研究才能满意地解答这一问题,这就意味着,研究者个人和单位所获得的荣誉就不会那么突显了。

20世纪90年代,英国海外发展管理局和世界卫生组织(WHO)最终公开资助了一项比较不同药物控制子痫惊厥

发作的研究。这项研究显示,最便宜的药比昂贵的药有效得多。试验报告有力地指出:“从硫酸镁首次推荐用于治疗子痫妇女(1906年)到安定的使用(1968年),大约3300万名妇女出现了一次惊厥发作,其中300万可能已经死亡。截至1987年苯妥英的引入,又有900万妇女可能出现了一次惊厥发作,其中100万死亡。”显然,由于研究方向的扭曲,人们为此付出了非常沉重的代价。

### 镁用于防治惊厥

“7年前,一项被评论家视为20世纪最重要的产科随机试验表明,在3种常用于控制子痫患者惊厥发作的治疗措施中,硫酸镁最为有效。从以下几个方面来看,这项合作性子痫试验具有里程碑意义:来自于9个发展中国家的27个医院的1687名女性患者及其医护人员参与了试验,这一样本量超过了过去50年内所有小规模非对照研究(大多在那些发病人数仅占全球子痫病例1%的国家开展)的样本量的总和。在英国这个硫酸镁并没有被产科医生广泛使用的国家,这篇试验的报告给临床实践带来了极大的影响。这就是发展中国家和发达国家合作共同来完善临床实践的很好的例子。”

Sheth S, Chalmers I. Magnesium for preventing and treating eclampsia: time for international action. *Lancet* 2002; 359: 1872-3.

有时,对现有研究进行系统评价将有助于暴露人们对广泛使用的但没有商业和学术价值的治疗措施效果的错误认识。有一篇系统评价探讨了因物理外伤引起脑损伤的患者使用类固醇激素将会提高还是降低患者的生存率。结果显示,无法确定类固醇激素对生存率的影响。过去几十年中,成千上万的患者接受了这一疗效并不确切的治疗措施,人们为此付出了沉重的代价。当最终完成了早就有必要开展的类固醇激素国际合作研究后发现,该治疗很可能导致了数以千计的脑损伤患者死亡。然而,该研究遭到了一些来自药厂和大学研究者的反对。为什么呢?这是因为他们参与了评价昂贵新药(即所谓的神经保护剂)的疗效的商业性试验,其结局指标可能对患者并不重要,所以他们不希望看到有其他研究来争夺试验参与者。

这一例子反映了解决那些对药厂和大学没有价值的问题的极端重要性,否则会导致患者利益受损。目前在英国,独立于药厂之外的临床试验普遍经费不足。英国国会下院健康特别委员会(House of Commons Select Committee on Health)已经意识到了这一点,英国临床研究协会(UK Clinical Research Collaboration, www.ukcrc.org)的创建以及英国医学研究理事会(Medical Research Council)、NHS研究与发展计划(NHS Research and Development Programme)和一些医学研究慈善机构制定的新策略都反映了人们对这一问题的认识。这些新举措是否有助于解答患者和医生关于治疗措施效果的疑问,这还有待时间的

考验。

要解答这些疑问的另一个原因是,需要确保宝贵的卫生资源没有被浪费。例如,20世纪40年代,静脉滴注人血白蛋白开始用于烧伤患者及其他危重患者的复苏。从理论上讲,白蛋白可以降低患者的病死率。令人惊讶的是,直到20世纪90年代才针对这一理论开展了公平试验。就这点而言,对相关对照试验进行系统评价没能找到证据证明人血白蛋白可以降低死亡风险。事实上,该系统评价提示,如果白蛋白对死亡风险确有影响,也是增加了死亡风险。

该系统评价的结果促使澳大利亚和新西兰的医生合作

开展了首个对人血白蛋白与生理盐水(另一种复苏液)进行对照的大规模公平试验。这项本该在半个世纪前就开展的研究显示,没有证据证明白蛋白的疗效优于生理盐水。而白蛋白的价格是生理盐水的20倍以上,因此过去50年间,全球价值约数10亿英镑的卫生资源就被这样浪费了。

#### 要点

- 许多研究质量低下且动机不纯;
- 药厂和大学都错误地引导着研究方向;
- 对患者非常重要的问题通常都没有得到解答。

## 第7章 进一步验证治疗措施 是每个人的职责

前几章已经讨论了开展低质量的或不必要的治疗措施效果研究会浪费掉多少时间、金钱和精力,这些研究现在不能,将来也无法解决与患者相关的问题。同时也总结了妨碍我们解决治疗性研究中不可避免的不确定性的一些因素。

第1章描述了某些新的治疗措施出现了意外的有害效应,而预期效应却未能实现。第2章指出许多常用的治疗措施和筛查试验没有经过充分的评价。第3章总结了公平验证治疗措施的基础,并强调了注重减少潜在偏倚和考虑机遇因素的重要性。第4章描述了部分几乎存在于医疗保健各个领域的大量不确定因素。第5章则比较了高质量、低质量和不必要研究之间的主要差异。第6章指出许多研究都是出于商业或学术目的,而未能解决可能影响到患者健康的实际问题。因此有理由相信,将来只有患者与研究者之间富有成效的合作才能更好地验证治疗措施的效果。

### 走出不确定性的迷雾

“只有当公众最终意识到现有可靠信息太少时,他们才会有动力积极参与到研究中。归根结底,要想改善临床医护效果和患者结局,就必需开展正确的研究,即有现实意义的研究,正如新成立的 James Lind Alliance 所倡导的一样。承认不确定性的存在并告知患者,是提高卫生保健质量、走出医学实践迷雾的重要策略。”

*Djulfegovic B. Lifting the fog of uncertainty from the practice of medicine. British Medical Journal 2004; 329: 1419-20.*

### 患者的选择:大卫与歌利亚

“谁能够判断研究问题是否真正满足了患者的最大需求?为什么最相关的问题没有被提出?目前是谁在设置问题?谁应该来设置问题?谁应该来指引大方向?患者最容易发现与其最相关的健康问题,也最清楚自己的健康状况、生活质量和生活数量。患者就像是牧羊人大卫,必须与歌利亚抗衡,歌利亚则代表了药厂和试验者,药厂需要获得研究证据以推销其产品谋取利润,而试验者一方面受到好奇心的驱使,一方面也需要保障其研究经费、学术地位和职业发展。商业利润、科学探讨、研究经费和学术论文都不能成为开展试验的正当动机,试验必须服务于患者利益。倡导高质量研究的患者和机构应准备好弹弓,选好石子,对准目标,然后射击。”

*Refractor. Patients' choice: David and Goliath. Lancet 2001; 358: 768.*

那么,患者怎样才能提高研究质量(包括研究内容及实施过程)呢?第6章提到了布里斯托的研究者如何与患者、风湿病学家、物理治疗师、全科医生共同合作,找出膝关节骨

性关节炎的研究内容与患者及其主治医生真正关心的问题之间的差异。临床医生和患者明显都对药物研究感到厌倦,他们希望看到的是对其他干预措施的评价,如理疗和外科手术、教育及应对措施。布里斯托的该项研究还表明,增进患者与医务人员之间的交流讨论将有助于了解他们真正的需求,他们最关心的有关治疗效果的不确定性是什么,以及研究应该解决哪些问题。

然而,患者和公众的想法是否真的能如数纳入研究呢?值得高兴的是,之前紧闭的医学世界大门终于向患者和公众敞开了,开始接受新的理念,以及所谓的“外行”,家长式的作风正在逐步消减。患者作为研究过程的参与者已受到越来越多的重视。来自问卷调查、研究报告的系统评价和试验报告的证据均表明,患者及公众的参与对于验证治疗措施功不可没。患者的经历可能引发深层次的思考并提供一些思路,他们的第一手资料对于了解人们对疾病的反应及其对治疗选择的影响很有意义。

变换角色有利于采取各种方式使患者和公众与医务人员共同合作,以促进医疗保健干预措施的验证,这将贯穿整个研究过程:

- 形成研究问题
- 方案设计,包括结局指标的筛选
- 研究项目的管理
- 制作提供给患者的信息资料
- 研究结果的分析与诠释
- 研究结果的传播与使用

为什么会这样呢?例如我们在第2章中提到,过度的治疗会给乳腺癌患者造成压力和自身的改变,这些压力和改变首先来自于临床医生和研究员,其次来自于患者本身。医生与患者合作共同确保研究证据既符合严谨的科学标准,又能够满足患者的需求。当乳腺癌患者对根治性乳房切除术提出质疑时,她们表示不仅关注癌症是否能够根除,还希望参与治疗方案的讨论。

如今,医生已不能将治疗和研究强加于患者,不再以患者无条件地依从治疗措施作为“成功”的衡量标准。最好能给患者提供他们认为重要且与其相关的治疗措施,鼓励患者共同参与治疗决策的制定,药物的选择,以及各种健康干预措施的验证。

### 谁来选择?

“对于参与决策(即谁来选择),大多数患者愿意共同协商而不是独立做出决策(共同决策模式),最讨厌医生自作主张(家长式领导模式)。”

*Thornton H, Edwards A, Elwyn G. Evolving the multiple roles of “patients” in health-care research: reflections after involvement in a trial of shared decision-making. Health Expectations 2003; 6: 189-97.*

### 患者需要信息

患者要在医疗保健过程中扮演合作者的角色,就必需获得高质量的易于理解的信息,无论他们是处在一对一的协商咨询关系中,或正在考虑是否参与一项试验,还是作为研究小组的成员。没有这些信息,就不会产生有意义的对话,也不能实现患者的真正参与。

患者常常会抱怨信息太少。尽管有些人不喜欢详细了解他们的病情和治疗方案,而愿意将所有的事都留给医生去考虑,但还是有很多人渴望获取更多的信息。他们想要知道和了解治疗措施的验证结果对他们本身会带来什么影响。对于他们而言,公开和透明是非常必要的,必须确保每个人都知道哪些验证治疗措施的试验正在开展;无论阳性或阴性结果都已公布;对所有相关证据进行了系统评价且保证了及时更新;不良反应没有被隐瞒。显然,患者需要确定研究者在启动新的研究之前已经清楚地了解了相关研究进展:正如我们在前几章中提到的,研究者有时懒得去了解相关的研究现状,患者因此受到伤害。

20世纪80年代,患者支持并参与了绒毛膜绒毛取样法的评价。绒毛膜绒毛取样是较羊膜穿刺更易早期诊断胎儿异常的一种方法。尽管这项新技术可能比羊膜穿刺更易引起流产,但所怀胎儿出现重度遗传性血液病风险极高(1/4)的孕妇仍然接受了这项检查。对于这类孕妇及其家属而言,尽管这项新技术可能会引起正常胎儿流产,但更要避免产下一个一出生就要遭受痛苦甚至夭折的婴儿。

### 消除不确定性

一位医学研究者因腰背痛做了血液检查,之后被告知部分结果不太明确,有些令人担忧。几年后,他被确诊为血癌(骨髓瘤)。这位研究者仔细回顾了整件事,其中就包括了证据的不可及性。

“有何经验教训?首先,作为一名患者,我强烈地感受到我作为一名医学研究者一直在努力奋斗的事业有多么重要,研究结果应向公众公开,以便人们做出事关自身健康的决定。令我不解的是,我没有办法对4项随机对照试验进行分析。为什么在得知了有相关信息存在但无法获取的情况下非要我做出决定呢?是因为研究结果不如预期的那样令人振奋?还是因为在不断发展的骨髓瘤研究领域出现了令人振奋的新理论(或药物)?我们到底还要容忍研究者的这种(前一朵花还没采完就又飞向另一朵花)的蝴蝶行为多久?遗憾的是,在这个临床研究主要被商业利益驱使的时代,这也不足为奇了。当你自己成了患者的时候,你就会质疑研究者为什么会忘了最基本的准则:研究首先应该是为了进一步验证各种假设,而非帮助厂商赢得市场。”

*Liberati A. An unfinished trip through uncertainties.  
British Medical Journal 2004; 328: 531.*

### 支持开展试验

长期以来,妇女们都在呼吁诸如超声扫描一类的新技术在常规用于孕妇之前应先对其安全性和有效性进行评估。为此,英国医学研究理事会启动了一项随机对照试验,从胎儿丢失率、母体副作用以及胎儿的短期和长期副作用等方面,权衡这类新技术的利与弊。

*Maternity Alliance Co-ordinator, cited in Chalmers I.  
Minimising harm and maximising benefit during innovation in  
healthcare: controlled or uncontrolled experimentation?  
Birth 1986; 13: 155-64.*

### 为什么我们需要开展大型研究来比较这两种技术?

“向孕妇提供有关某检查技术利弊的准确信息非常重要,为了做到这一点,我们需要在不受其他任何因素影响的条件下比较这两种技术。唯一的办法就是让一半患者接受其中一种检查,另一半接受另一种检查,并且随机决定哪些患者进入哪组。

一开始,很多人认为随机决定患者的分组情况有些不恰当,但事实上这是最科学的方法,因为这能避免因其他各种因素可能产生的偏倚。目前,我们任何一个人都不知道哪种检查技术更好,每种技术都有利有弊,对每种技术的支持和反对意见也都差不多。如果您以及其他志同道合的人愿意加入,我们很快就能进一步了解这两种技术。很多医院都已经参与了这项研究,我们将尽快找出相对更好的那项技术。”

*From the patient information leaflet developed by patients'  
organizations for the MRC amniocentesis  
versus chorion villus sampling trial.*

然而,对于其他妇女来说这种权衡是很困难的,她们怀有异常胎儿的可能性不太大,可能就是1/50~1/200。对她们而言,更重要的是要知道绒毛膜绒毛取样是否比羊膜穿刺更易导致流产或其他问题。因此,英国医学研究理事会支持一个国际合作机构着手研究这些问题,这一倡议得到了新闻媒体的广泛支持。例如在英国,许多新闻工作者都赞成这项研究,并强调在推广新的医学技术之前必须先审慎评估。

一位临床医生在写给《卫报》(The Guardian)的信中给出了一个极不负责任的建议:他建议妇女直接去接受这种虽新却未经试验充分验证的技术。这一建议引来了孕妇联盟(由多个压力团体组成的伞型组织)一名协调员的强烈反对。这名协调员强烈要求开展研究验证这一新技术。

在当时,人们对这样一项随机对照试验的推进可谓前所未有的,一个名为改善孕妇服务联盟的民间团体召集了多个志愿者组织和患者群体,开会讨论并鼓励他们公开支持英国医学研究理事会的研究计划。各团体的代表参与起草了

一份提供给所有试验潜在参与者的传单,其中的信息明确表明已有7个民间团体正式公开支持这一试验。此外,这些民间团体还一致认为,试验组织者应竭力确保所有孕妇暂不接受这项新技术(除外试验参与者),除非她曾经产下过患有遗传性疾病的婴儿。

早期活跃在乳腺癌研究领域的人们也意识到需要获得准确的信息以认清并挑战现状。首先,他们进行了自我教育,以确保自己采取的行动是行之有效的,然后再去教育他人。例如,在20世纪70年代早期,美国的一名乳腺癌患者同时也是一位自由作家的Rose Kushner,敢于向当时传统的独裁式医患关系及根治术的必要性发出挑战。她在对根治性乳房切除术有效性的相关证据进行了全面评价后写了一本书。到70年代末,其观点已被公众广为接受,美国国家癌症研究所也邀请她共同审核新研究的试验方案,影响甚广。

20世纪70年代英国的Betty Westgate被诊断为乳腺癌的不幸遭遇促使她成立了乳腺癌切除术协会。这便是乳腺癌诊疗中心的前身,如今已是一个在英格兰、苏格兰和威尔士拥有多处分支机构的日益发展壮大的组织。乳腺癌诊疗中心帮助成千上万的女性患者获得信息与支持。另一名癌症患者Vicky Clement-Jones设立了CancerBACUP作为国家慈善机构,该机构不仅给患者提供实用的建议和情感支持,也提供高质量的治疗和研究信息。如今,CancerBACUP及其癌症专业护理人员每年为近5万名患者提供这样的服务。

20世纪80年代晚期,美国HIV感染者或艾滋病患者是另一个消息灵通且组织严密的特殊群体。他们敢于捍卫自

己的权益,为患者参与试验设计铺平道路。最终试验提供的治疗方案更具多样性可供患者选择,试验设计也更加灵活,鼓励了更多的患者参与其中。同样,20世纪90年代早期,英国的一组艾滋病患者参与了伦敦切尔西和威斯敏斯特医院的研究,这些患者帮助设计了试验方案。

这些艾滋病领域的积极分子使得试验者不得不反思:部分研究者将组织有序的患者团体视为祸害,但事实上这也是对试验者解决治疗措施的不确定性提出了合理的挑战。在此之前,研究者的做法往往忽略了患者所关注的结局。另一方面,患者开始意识到对新药疗效草率做出判断的危险性,以及在严格审评前就将所谓的“大有希望”的艾滋病新药投入使用的危险性。研究者可能会争辩,这种出于同情而给患者使用了新药的做法只会延长目前和未来患者继续忍受不确定性的痛苦。然而,患者却认为这最终会促使患者和研究者意识到,有必要在试验中对治疗措施进行充分的对照评价,试验设计应由医患双方共同完成,且应同时兼顾双方的需求。

20世纪90年代,一项艾滋病试验特别强调了患者参与的重要性。当时正值齐多夫定(zidovudine)刚开始用于治疗艾滋病患者。在疾病晚期患者中,已有高质量证据证实了该药的疗效。但接下来的问题是,齐多夫定在感染早期是否能够延迟疾病进展并进一步提高生存率。因此,美国和欧洲均开展了试验以验证这种可能性。美国的试验在发现了可能存在但仍不确定的疗效后提前终止了。尽管美国试验已经有了结果,但在患者的积极参与以及患者代表的一致认同下,欧洲试验一直持续至有明确的终点出现。结果却与美国试验大不相同:齐多夫定用于感染早期并不会产生任何疗效。在这种用药环境下,唯一能够明确的竟是该药的副作用。

遗憾的是,HIV或艾滋病领域的经验教训并没能引起有类似困境的其他领域的重视。由蛋白酶传染性因子引起的罕见疾病(如疯牛病)是年轻人都难以承受的,往往会迅速致命。患者家属不顾一切地要去试用某种新药的做法事实上可能会延误人们发现真正利大于弊的治疗措施。一名患有蛋白酶传染性因子疾病的年轻患者的父亲在网络了解到一种从来没在人体身上用过的药物,日本的啮齿目动物实验提示该药可能对蛋白酶传染性因子疾病有效。因为该药尚未获准用于治疗人蛋白酶传染性因子疾病(必须采用颅内直接给药,因而可导致出血),所以这位绝望的父亲只好向高级法院提出申请要求将这种药物给他儿子使用。法官的

#### 共同影响未来研究

“癌症研究必须充分考虑所要帮助对象的需求及利益。癌症专家通常想要解决他们对癌症诊疗认识的不足,但患者及其亲友可能认为其他方面的癌症医护更值得继续研究。”

*Cancer BACUP. Understanding cancer research trials (clinical trials). London: Cancer BACUP, 2003.*

#### 非专业人士共同反思艾滋病

“艾滋病领域的可信度斗争是多边的:斗争者竟然来自各个不同的领域。非专业人士参与宣传和评估科学观点帮助人们认识到了什么是艾滋病,也让我们难以分辨谁是非专业人士谁又是专家。每当危急时刻来临,我们就会质疑到底是特定的知识论点还是某人说的话更可信。但从更深层次上理解,这涉及到可信度的评价机制:如何判定科学观点的可信度?由谁来判定?争辩科学观点的同时又在辩论着怎样做或由谁来做才算科学。”

*Epstein S. Impure science: AIDS, activism and the politics of knowledge. London: University of California Press, 1996.*

#### 研究者支持患者参与试验设计

“我们鼓励患者和患者团体参与临床试验的设计。这能够进一步确保试验目的和试验设计得以准确实现,也能够让艾滋病患者更加珍惜参与试验的机会。”

*Byar DP, Schoenfeld DA, Green SB, et al. Design considerations for AIDS trials. New England Journal of Medicine 1990; 323: 1343-8.*

结论是：“虽然治疗本身不应视为一项研究，但这也是初次了解 PPS（多硫酸戊聚糖，尚未上市的药物）用于 vCJD（变异型克雅氏病，蛋白酶传染性因子疾病的一种）患者效果的一个机会。”法官的言辞说明她对此事的认识不够，这令人担忧。她可能没有意识到不设对照的试验可能会延误人们发现真正对蛋白酶传染性因子疾病有效的治疗措施。她可以把判决建立在这样一个条件之上，即必须详细记录这名患者以及后续向高级法院提出类似申请的患者的治疗情况和病情进展。如果她这么做，我们就不会像现在这样对这种未经评价的治疗措施的效果一无所知了。

近期，一名早期乳腺癌患者质疑 NHS 为什么决定不给她使用新药赫赛汀（群司珠单抗，trastuzumab）。英国卫生大臣介入了此事，NHS 最终妥协。赫赛汀可能对晚期乳腺癌患者有效，但同时也可能导致心力衰竭。目前尚未在疾病早期对该药进行过充分验证，最终结果可能会是弊大于利。但现在还言之过早。

### 妨碍治疗措施的公平验证

让患者参与研究显然不是万能的。举个例子，研究者的一项调查显示，虽然让患者参与临床试验效果很好，但同时也暴露了一些很现实的问题。许多问题可能是因患者对于研究的具体程序和资金来源认识不够所致。

我们编写此书的目的之一就是为了让人们进一步了解验证治疗措施的原则以及有关治疗措施效果的可靠证据是如何获得的。通过这种方式，我们希望帮助那些愿意共同提高证据质量的患者采取更多行之有效的行动。正如我们在第 3 章中所强调的，让患者意识到偏倚和机遇因素的误导作用非常重要。

如果患者对研究的基本原则不够了解或不够重视，有时可能反而会妨碍治疗措施的公平验证。就像那名 vCJD 患者的父亲那样，有时会在绝望的情况下不顾一切地去争取那些尚未被充分验证的治疗措施，这些治疗对于濒临死亡的患者都可能弊大于利。患者或患者团体要求医疗机构出于同情而给艾滋病患者使用“大有前途”的新药，这种做法是不妥的：它延误了人们发现可直接改善患者结局的治疗措施。

这样的例子还不止一个。20 世纪 90 年代中期，干扰素开始用于治疗复发-缓解型多发性硬化，而当时相关的疗效证据非常匮乏。2001 年，研究者对这类药物的临床对照试验进行了系统评价，所有试验均由药厂组织完成。结果表明，虽然干扰素使多发性硬化的复发率略有下降，但没有证据证明它能够延缓残疾的持续进展，比如需要使用行走辅助工具或轮椅。

鉴于每位使用干扰素的患者年治疗费用超过 1 万英镑，英国国家卫生和临床医疗优选研究所（NICE，一个服务于 NHS 专门评价治疗措施的成本效果的机构）分析后认为：使用这类药物或另一种名为格拉默的药物，实际上是对有限资源的浪费。许多患有这种消耗性疾病的患者和患者组织对

此非常震怒。他们对 NHS 拒绝提供还有一线希望的药物感到气愤，然而他们是否清楚地意识到现有的证据还远远不能让人信服？相关研究只公布了部分结果，所用结局指标的相关性令人质疑，随访期也只有 2~3 年，而这种疾病至少要持续 20 年。

在这种压力之下，政府也只好妥协。英国卫生部和药厂共同制定了一个 10 年计划向多发性硬化患者提供干扰素。但这却让人们无法了解干扰素是否对患者有效。更糟糕的是，因为干扰素现已成为公认的治疗这种疾病的标准药物，从而导致针对此病的其他治疗措施的研究突然变得十分昂贵。这是为什么？因为医生和患者一致认为将其他新的治疗措施与安慰剂对照违背伦理，必须将新的治疗措施与昂贵的干扰素进行对照。

### 评价多发性硬化治疗药物

1. NICE 宣布 NHS 不推荐将  $\beta$ -干扰素或格拉默用于多发性硬化的治疗。
2. 英国卫生部计划通过“风险分担计划”向患者提供这类药物，但该计划既不科学也不切实际。
3. 随机试验表明硫唑嘌呤（比干扰素便宜 20 倍）可能与干扰素有同样的效果。
4. 这类药物的长期疗效尚不明确。
5. 政府经费更应用来开展一项长期随机试验，将  $\beta$ -干扰素或格拉默与硫唑嘌呤和安慰剂进行对照。

*Sudlow CLM, Counsell CE. Problems with UK government's risk sharing scheme for assessing drugs for multiple sclerosis. British Medical Journal 2003; 326: 338-92.*

药厂资助的试验在各方面都相对完善，但唯独缺少独立的学术评价，从而导致有关多发性硬化的治疗进展受阻，无论用什么指标来衡量结果都不能令人满意。

这个令人遗憾的故事还有另外一个重要方面，那就是经济利益关系在很大程度上阻碍了对干扰素治疗多发性硬化的合理验证。药厂销售干扰素，神经科医生和其他临床医生推广该药则会从药厂处获得利益。所以有些神经科医生使用昂贵的脑扫描对患者进行检查，然而这一切仅仅是基于尚未被证实的假设：脑扫描图是监测疾病进展的有效途径之一。厂商对这种检查手段极为推崇。尽管鲜有证据证明脑扫描可有效监测多发性硬化的病情进展，许多神经科医生还是坚持认为这种检查方法必须用于针对该病治疗的所有临床试验中。虽然药厂可以为受试者支付这些尚未被确证的检查方法的高昂费用（每次扫描的费用高达 2 000 英镑），但对于那些希望不借助药厂的力量独立完成研究的人而言，这笔费用将是个大问题。

另一个相对不那么明显的利益冲突存在于患者组织与药厂之间。许多患者组织都是由志愿者组成的，资金匮乏。与药厂合作或接受药厂的资助能帮助这些组织发展壮大，扩

大其影响力,但同时也可能会歪曲患者的意愿,不能代表患者的利益,如改变研究方向。有时患者组织是由药厂组建的,目的是为了帮助推广其产品。举个例子,一家生产干扰素的药厂(后来受到英国药品管理局的指责)组建了一个名为“诉求行动”的新患者团体,敦促 NHS 向多发性硬化患者提供干扰素。患者团体了解到的信息都是干扰素的确有效只是很昂贵,事实上关于干扰素是否有效当时还是一个未知数,这种错误的观点持续了一个世纪之久。即便到现在,时隔开展首项干扰素试验 16 年之后,依然没有有效的长期数据证明干扰素有效。

我们不仅高度怀疑这些治疗多发性硬化的新药是否真的对患者有效,我们还应警惕有些药物可能引发非常严重的副作用。在美国,药品食品监督管理局仅基于 12 个月的研究数据就批准了生物药 Tysabri(那他珠单抗)。这一决定所带来的影响还好没有持续太长时间:在为数不多的参与了该药临床试验的患者中,出现了两例非常罕见的致命性脑病,药厂因此撤回了该药。尽管这场悲剧几乎可以肯定是这种新药造成的,如果暴露时间更长结果可能会更糟,但患者和医生还在为此争吵不已。

### 进一步验证治疗措施的步骤

上文指出了患者参与验证治疗措施可能引发的问题,以及他们通过何种途径可能无意干扰到治疗措施的公平验证。

#### 慈善机构和患者团体应当声明利益冲突

“药厂不可能无私地为慈善机构捐款。在欧洲不允许直接向患者宣传处方药广告,但众所周知患者团体的游说能力足以影响政府和 NHS。”

“患者信任医疗慈善机构,希望他们提供的信息是没有偏倚的,是不受其资助机构的影响的。接受了药厂资助的慈善机构应该对此做出利益冲突的声明,提醒患者质疑该机构及其所提供信息的独立性,并从其他自主性更强的机构进一步获取信息。”

“一个关节炎患者团体在获取了带有误导性的阳性结果后(本来是 12 个月的试验,但报告的却是 6 个月的结果),发起了一项要求政府和医生更多地使用新环氧化酶 2 抑制剂的活动。他们并没有声明这次活动是由药厂 Pharmacia 公司和辉瑞公司资助的。”

“性无能协会在辉瑞公司的资助下开展了伟哥的推广活动,并将辉瑞的标志挂在了该协会的网站。英国糖尿病协会接受了 11 个生产糖尿病药物的药厂近一百万英镑的资助,但在其年度报告中并未提及此事。”

“这些组织没有公布其获得的资助金额,这不得不引起人们的怀疑:为什么不能大胆地向公众公布这些数字呢?这里面到底有多少不为人知的秘密?”

Hirst J, Charities and patient groups should declare interests.

British Medical Journal 2003; 326: 1211.

与其他事情一样,好心不一定办得成好事。无论怎样,还是有很多患者与研究者共同合作改善了研究设计及其临床相关性的成功案例。因此,许多研究者都在积极地挑选能够与他们合作的患者。

旨在探讨如何在临床实践中改善共同决策和风险沟通的研究应该由患者和研究者共同完成。曾经有这么一项试验就是研究者和患者一起合作制定的试验计划。首先,他们查阅了相关文献以确定在患者向医生咨询的过程中,医生认为哪一方面对患者来说最重要。然后他们通过组织多个患者和居民团体,直接调查患者的观点。他们探讨了临床决策的模式,信息的可及性以及患者对参与决策的态度。他们采用了严格的研究方法以确定医生和患者的观点存在哪些共同之处。

被调查的患者指出许多研究者之前提出的以及文献报道中使用的结局的确很重要,比如患者知道自己可以参与决策以及医患双方达成共识。然而,被调查者还指出了一些他们认为重要的结局,比如决策过程中当他们提出了有意义的建议时会受到尊重。他们提出患者应该能够从其他渠道获取更多的信息以帮助其评价研究者提供的信息。被调查者希望能有机会让其他人也参与其中,比如患者家属、其他专家、相关的医务人员(护士和咨询医生)、自发或自助团体。他们强烈要求确保医护的持续性,包括明确的管理计划和安排定期审核治疗决策,这一切都应基于总结材料或录音信息。

举个例子来说明研究者与患者共同完成试验准备工作的意义。一项卒中试验的研究者与患者以及潜在的患者共同探讨急诊环境下验证治疗措施所遇到的难题。要想急性卒中治疗获得成功,就需要在卒中发作后及早开始治疗。因为研究者不确定采用何种方法开始治疗才是最有效的,因此他们征求了患者和护理人员的意见。研究者召集了一组患者和医务人员一起参与讨论会,并且组织了由老年人组成的焦点团体。会上研究者阐明了试验计划,患者则帮助研究者起草并修订了涉及试验信息的宣传资料。

#### 研究者与患者合作,实现互利互惠

政府资助的 PRISM 试验是一项在英国开展的多中心随机对照试验,旨在比较针对骨 Paget 病的各种治疗方案。英国 Paget 病协会(NARPD)是英国唯一一个支持 Paget 病患者的组织,该协会从一开始就与 PRISM 研究小组有着密切的合作。NARPD 的参与是整个试验实施过程中不可或缺的一部分,并且在其中扮演了具体的角色:同行评审、试验指导委员会成员、为试验参与者提供建议以及在 Paget 病患者中宣传该试验。

摘自 Langston A, McCallum M, Campbell M, Robertson C, Ralston S.

An integrated approach to consumer representation and involvement in a multicentre randomized controlled trial.

Clinical Trials 2005; 2: 80-7

这项全面的预试验使得研究伦理委员会很快就批准了相应的随机对照试验方案。焦点团体的参与者也意识到了要从急性病患者那里获得知情同意对于研究者而言的确是个伦理困境,如果患者意识清醒,这样做反而可能让患者感到更困惑。由于有了患者的参与和建议,试验设计方案最终能被各方都接受,涉及试验信息的宣传资料也大大改进。

### 患者参与研究策划的重要性

研究者召集了几个老年团体共同讨论一项关于如何治疗急性卒中的新研究的设计。当卒中真正发作时,患者通常都无法表达自己的观点了。研究者得出了以下几点结论:

- 让患者参与卒中试验的设计是很有意义的;
- 来自于未发生过卒中的人以及卒中患者的护理人员意见能够大大改进涉及试验信息的宣传资料;
- 人们赞同根据患者的临床情况采用不同的知情同意方法;
- 在新的随机对照试验策划过程中,患者的参与是非常重要的组成部分。

摘自 Koops L, Lindley RI.

*Thrombolysis for acute ischaemic stroke: consumer involvement in design of new randomized controlled trial. British Medical Journal 2002; 325: 415-7.*

第6章阐述了英国一项局灶性前列腺癌试验中患者参与试验准备工作的重要性。研究者有充分的理由去比较积极监测相对于手术治疗或放疗的效果,从而消除有关这些治疗措施的不确定因素。显然,这是几种完全不同的治疗形式,那么患者和医生会怎样面对试验所采用的随机分配方案?研究者不难想象医生在说服患者参与试验的过程中可能遇到的种种困难。他们也不难想象患者将会难以决定是否要参与试验,尤其是当患者需要同意被随机分配到三个完全不同的治疗组中,并且不同的治疗组都有着不同的潜在并发症风险。

在这种情况下,研究者制定了一个由两阶段组成的研究计划:首先开展可行性研究,然后再基于结果实施主体试验。第一阶段的试验结果证明试验是可行的,大多数男性患者都愿意参与这项由三个治疗组(积极监测,手术治疗或放疗)组成的试验。

### 患者和公众应如何参与?

患者和公众可以通过多种途径参与治疗措施的验证。如上文所述,他们可能会是研究的源动力,可能是他们首先发现认识缺口以及对新治疗手段的需求。他们的参与可能是受到了研究者的鼓励;他们可能参与研究工作的某些阶段而并非全过程;他们可能参与确定哪些具体的不确定性值得研究,生产并更新系统评价以传播并合并研究结果;即便对于同一项研究,他们也可以通过不同的方式参与其中。有时他们自己就启动了研究工作。这里并没有任何硬性规

定:研究针对患者参与所采用的不同策略和途径的恰当性取决于研究本身。正如局灶性前列腺癌试验和患者参与决策试验所阐明的那样,患者参与研究的途径是不断变化的,甚至在一项研究的不同阶段都可能采取不同的方式。

2004年发表了一篇有关患者参与的系统评价,此文探讨了患者参与在发现和优选研究主题方面的作用。作者最后得出结论:现在那些研究计划的负责人已经有了丰富的经验通过直接或间接地与公众合作让患者参与研究的设计与实施。如今,研究的资助方通常会要求研究申请者明确他们将如何让公众和患者参与其研究计划,也会建议研究者通过某些途径去进一步了解如何做好这一点。

### 针对患者参与研究的管理规范指南

1. 患者所扮演的角色应由研究者和参与研究的患者共同确定。
2. 研究者应对患者参与研究所需的费用做出合理的预算。
3. 研究者应尊重不同患者在技能、知识及经验上的差异。
4. 应向患者提供培训和个人支持,以帮助其参与研究。
5. 研究者应确保自己掌握了必要的技巧能够顺利地让患者参与到研究过程中。
6. 患者应参与决定如何招募研究的受试者,并且随时有权了解研究的进程。
7. 研究报告应对患者参与有所描述。
8. 应以患者容易理解的形式和语言向患者公布研究结果。

摘自 Telford R, Boote JD, Cooper CL.

*What does it mean to involve consumers successfully in NHS research?*

*A consensus study. Health Expectations 2004; 7: 209-20.*

总之,患者和公众参与研究的时代已经来临。他们不再被忽略,也不会仅仅是做做表面文章。作为研究的合作者,他们可以优选研究主题,提高研究质量。虽然患者有时可能会无意干扰到治疗措施的公平验证,但患者与研究者的共同合作能够减少有关治疗措施的不确定性,实现互利互惠。

### 要点

- 患者与研究者合作能够减少有关治疗措施的不确定性,实现互利互惠;
- 患者能够影响针对治疗措施疗效的研究;
- 患者的参与能够提高研究质量;
- 患者有时会无意干扰到治疗措施的公平验证;
- 应该让患者能够很容易地全面获取试验信息,这样才能使患者的参与更有意义。

## 第 8 章 改革蓝图

医学研究对提高生活质量和延长寿命的贡献毋庸置疑,然而我们在这本书中阐述了当今科研的“驱动者”(商业和学术利益)在考虑患者所关注问题方面还做得远远不够。医疗保健的过程总是充满了不确定性,无论其重要程度的高低,如何找出比 NHS 更好的解决办法,与我们每个人的利益都休戚相关。那么,如何引发一场改革,在这场改革中我们不会回避日常医疗保健中存在的确定性,而会将其视为一种挑战;同时,验证治疗措施的研究也将在这场改革中成为优质医疗保健的重要组成部分。

在勾画这场改革蓝图之前,我们先举例让大家了解我们希望未来的日常医疗实践会发生什么样的变化。

### 琼斯先生的故事

琼斯先生(Ifor Jones)是威尔士北部农村的一位退休农民,由于常常感到疲惫,于是他决定去看全科医生。琼斯先问了女儿何时能开车送他到 10 里外的诊所,然后预约了医生。医生和琼斯一起讨论了症状,并给他做了检查,然后抽血送到当地医院化验。几天后,化验结果出来了,显示琼斯患了一种因缺乏维生素 B<sub>12</sub> 导致的贫血(恶性贫血,参见第 4 章)。全科医生立即打电话给琼斯让他再来一趟。

于是琼斯又去了。医生告诉琼斯血液检查证实他患有恶性贫血,注射维生素 B<sub>12</sub> 对这种贫血的效果很好,但医生同时指出,专家对注射频率问题意见不一,即只需每三个月注射一次还是需要更频繁的注射。医生在网上联系了国家卫生图书馆,然后给琼斯看了一篇比较不同注射频率的临床对照试验的系统评价。该系统评价证实没有人知道更频繁的注射是否比每季度注射一次更能有效地缓解疲劳。

这个网站同时也链接了一些患者界面友好信息,提示 NHS 研究与发展项目组正在开展一项临床对照试验,比较以不同的频率注射维生素 B<sub>12</sub> 对于恶性贫血的治疗效果。发起这项研究是因为几个全科医生和部分患者通过临床问题解答服务系统指出,他们想知道恶性贫血患者应多久接受一次注射。鉴于每个月频繁地到诊所接受注射会很麻烦,但可能更有效地改善琼斯的症状,因此医生问琼斯是否愿意参与这项试验。

当研究者和患者设计这项试验以比较不同注射频率的效果时,患者明确指出除了要进行恶性贫血的标准血液化验外,还必须评价最令患者烦恼的症状(如疲劳)。琼斯和他的医生认为参与这项试验非常容易:除了医生会常规搜集有关治疗效果的监测信息外,不再需要别的信息了。而且,这项试验的初步结果预计会在一年内得出,这就意味着琼斯本人有望受益于这一新证据。

医生问琼斯是否愿意把试验提供给患者的资料带

回家,以便认真考虑是否要参与这项试验,而琼斯当场就决定要参与。医生在这项试验网站中受密码保护的保密区内输入了琼斯的详细资料,几秒钟后就收到了试验协调中心的信息,提示琼斯被随机分配到了每三个月注射一次的试验组。

此后每季度,试验协调中心都会发短信给琼斯及其医生,让他们回答一些有关琼斯的症状和体征的问题,并提醒他们将血液样本送到实验室。然后,实验室会将化验结果抄送给试验协调中心、琼斯和他的医生。

一年刚过,医生和琼斯就收到了试验的初步结果。他们综合考虑了这些结果,以此来决定是继续每三个月注射一次还是选择更频繁的注射。就这样,为了减少这个对琼斯及其医生都很重要的问题的不确定性,他们都贡献了自己的力量。

上述例子反映了如何服务于维生素 B<sub>12</sub> 缺乏性贫血患者及无法确定不同治疗方案的相对优势的全科医生。尽管早在一个世纪前就发现了针对此类贫血的有效治疗措施,但关于维生素 B<sub>12</sub> 注射频率的问题一直未得到解答,这是因为该问题对药厂或研究者都没有多大好处。只有深入基层了解患者和医生的实际情况,这类问题才会浮出水面。就这个事例而言,除了常规的医疗实践外,无需作其它努力便可得出答案。

我们以令人痛苦的慢性病的疗效不确定性为例,阐述了这类问题的解决办法。其实,解决任何医疗不确定性问题都应遵循这种基本方法——从危及生命的紧急抢救(如急性中风)到令人烦恼的自限性疾病(如普通感冒)。那么,我们每个人能做些什么来确保上述方法被 NHS 广泛采纳呢?

### 我们的蓝图

光看一点,下面没有一条建议称得上是革命性的,甚至谈不上新颖。但综合来看,只要患者与医生联合起来共同推进,下列 7 条行动倡议就会构建出一幅关于治疗措施的使用与验证的改革蓝图。

(1) 当疗效不确定时,鼓励医生和患者诚实面对。通常,要让医生承认自己不能确定某治疗措施的效果是一件很难的事,患者有时也不愿接受。因此,有时医生会给患者一种虚假的安全感。但如果医生和患者要想共同合作成功地完成疗效评价,那么双方就必须承认这样一个事实,即如果治疗措施未经充分评价,则可能带来巨大危害。此外,双方还必须熟悉获得可靠证据所需要掌握的方法。为了让这一切成为可能,我们必须找到促成医生与患者合作的最佳途径。

(2) 在决定是否认同临床试验和非临床试验提供的治疗措施时,医生常面临双重标准。对于那些愿意承认疗效的不确定性并准备通过正规的治疗对照来解决这种不确定性的医生,与患者沟通比与其同事沟通更需要遵循严格的规则。当不确定某治疗措施的效果时,标准的解决办法就是开展对

照试验或采用其他无偏倚的评价方法来评估疗效。我们应确保参加疗效评估研究不是一种冒险的行为,也就是说所谓“标准”的治疗方法通常都安全有效。

(3) 学会如何判断是否应相信治疗措施所声称的疗效。我们需要更广泛的公众参与,共同避免偏倚和机遇因素对疗效评估证据的严重歪曲。科学研究的一大特点是发现偏倚并使偏倚最小化,然而很难说这一点已经成为了“常识”。我们需要下定决心来减少这种认识上的重要差距,并把这些理念作为常规的教育内容从学龄期的学生就开始灌输。

(4) 提高生产、保存并传播有关疗效研究证据的系统评价的能力。很多有关疗效的迫切问题的答案很容易就能找到,即通过系统评价现有证据、更新这类系统评价、并将评价结果迅速传递给医生和患者。但在制作系统评价的过程中要想轻松获取已有研究证据的信息,还有很长的路要走。我们应督促 NHS 把这个作为其主要目标,这样才能合成有关疗效的可靠信息,并且在整个医疗服务过程中可轻松获取。

(5) 清理临床研究领域中的学术腐败和利益冲突。很多人惊讶地发现,研究者在其新研究寻求伦理审批和资金资助时,竟然不需要对已知证据进行系统评价。这样做难免会造成一些设计差、甚至完全不必要的研究仍继续开展,这在伦理或科学上都是不能接受的。我们应敦促研究资助者和研究伦理委员确保研究者在没有参考其他相关证据的系统评价之前不得启动任何类型的新研究。此外,有偏倚的漏报或瞒报也是不合伦理的。必须要求研究者公布患者参与的所有研究的结果,并说明这一新证据对相关领域现有证据总的意义。

(6) 要求药厂提供更好、更全面、更相关的疗效证据。2005年,英国下议院健康委员会公布了一篇关于药厂影响力的惊人报道。过去,只要药厂的势力和影响力受到质疑,药业巨头们就以终止其在英国的研发活动这种近乎威胁的方式来做出回应,并且发出警告称此举将对英国经济产生可怕的后果。连续几届政府都听到了这些声音并且妥协了,没有采取任何措施去遏制药厂的这种蛮横行为。健康委员会的建议必须引起高度重视。我们尤其应该要求所有临床试验从一开始就公开注册,试验一结束就应公布所有试验结果。

(7) 确定并优先研究患者和医生都认为重要的疗效问题。那些在可预见的将来不大可能使患者受益的基础研究以及那些服务于药厂利益的研究占据了研究资助单位资助项目和学术机构研究活动的绝大部分。那些不可能赚钱但对患者非常重要的应用研究,纵然得到公众支持,也必须想尽一切办法才有立足之地。我们应该关注 NHS 是否做了更多工作以确定患者和医生关心哪些疗效问题,同时还应关注研究资助者在优选研究时是否考虑了这些问题以减少其不确定性。

### 行动计划:您可以做到的事

确定对您很重要的疗效问题。

检索国家卫生图书馆([www.library.nhs.uk](http://www.library.nhs.uk)),查看是否有来自最新系统评价的可靠信息。

如果找不到您需要的信息,请登录 NHS Direct 官方网站或拨打热线电话[[www.nhsdirect.nhs.uk](http://www.nhsdirect.nhs.uk); Tel: 0845 46 47 (英国或威尔士); 08454 24 24 24 (苏格兰)]。专家也许能帮您找到这些信息。

如果找不到可靠的信息来回答您的问题,请咨询 NHS Direct 是否有人正在对现有相关证据进行系统评价,或是否有可能相关的临床对照试验正在开展。

如果的确有相关试验正在开展,请登录 [www.controlled-trials.com](http://www.controlled-trials.com) 或专门的临床试验网站来了解您是否可以参与其中,以及您可以和谁来讨论您的选择。

在同意参与某临床试验之前,请您先确认是否满足下述条件:(i) 试验方案已在 [www.controlled-trials.com](http://www.controlled-trials.com) 上公开注册;(ii) 试验方案提及,现有证据的系统评价显示确实应该开展该试验;(iii) 您获得了一份书面承诺书,保证所有研究结果都将公布并送到那些表示愿意收到试验结果的所有参与者手中。

如果您的问题没有得到充分解答,也没有人对此感兴趣,请将问题呈交给下述机构供其审议: NHS 卫生技术评估 (HTA) 项目 ([www.nchta.org](http://www.nchta.org))、英国国家卫生和临床医疗优选研究所 (NICE) ([www.nice.org.uk](http://www.nice.org.uk)) 以及相关的医学研究慈善机构 ([www.amrc.org.uk](http://www.amrc.org.uk))。

学会发现问题、提出问题并寻求可靠的答案。

如果您想要讨论有关疗效的证据以及存在哪些不确定性,请让您的医生知道您的想法。网站 [www.ohri.ca/DecisionAid](http://www.ohri.ca/DecisionAid) 上的信息也许会对您有所帮助。

让更多的人了解偏倚和机遇因素的影响,说服相关人士将这种教育从小学开始就纳入课程。

对于毫无根据的厂家宣传和媒体报道,所谓的某治疗出现了“重大突破”,要持有怀疑态度;应提出中肯的问题,并与您的朋友讨论。

对于那些仅基于信念和教条就提供给您或您家人的治疗措施,若没有经过可靠证据的证实,都应提出质疑。

鼓励与医生、研究者、研究资助者及其他相关人士合作,共同推进研究,解决那些尚未得到充分解答而您又认为很重要的问题。

### 结束语

早就应该发起验证治疗措施的革命了。如果医生和患者合作,我们所提倡的方法将非常可行。让我们立即行动起来,共同改变现状!

杜亮,蔡羽嘉,文进,杨晓妍,陈敏,刘婷兰,陈耀龙,原志芳,  
敬媛媛,孟月,毛秀,周邦旻,喻佳洁,宋莹,李胜吾 译  
刁 驥,张鸣明,李幼平 审校