

第二章

治疗与循证医学

本章要点

- 应用 PICO 模式能梳理出临床问题精炼的构架,是实践循证决策的第一步。
- 治疗性研究科学性评价最重要的方面是是否采用随机、双盲对照的设计方法。
- 在具体实施治疗性措施循证医学时,还需关注治疗性措施的可获得性、可行性、经济负担、患者的意愿。

教学目的

- **掌握:** ① 随机、双盲、对照设计方法;② 实施治疗性措施循证医学步骤。
- **熟悉:** 治疗性研究常用的设计方法及优缺点。
- **了解:** 治疗性研究疗效大小的评估。

临床治疗疾病的目的是:① 治愈疾病或稳定病情;② 预防疾病的复发和防治并发症;③ 缓解症状、改善脏器的功能状态,同时提高患者的生命质量。如何选择权衡治疗措施的利和弊,如何获得最可靠的证据,以指导临床工作者对具体患者提供最好的治疗,这些都是每个医师在日常临床工作中经常碰到的问题。因此,需要了解有哪些治疗性研究,它们在评价干预措施时的优缺点;在获得治疗性研究后如何评价它的科学性,治疗的疗效是否足够大,在将治疗措施应用于患者时需要关注哪些方面。

第一节 应用 PICO 模式提出临床问题

针对解决临床问题进行治疗性研究的文献检索或者研究方案设计时,首先均需明确关键的临床问题是什么?应用 PICO 模式能梳理出临床问题精炼的构架,是循证决策的第一步。以新型口服药物治疗类风湿关节炎的疗效和安全性评估为例。

P(patient): 研究对象,首先确定患者的一般特征如性别、年龄区间等,其次是患者的疾病特征包括病情轻型或重型、疾病活动或非活动、疾病分期、病程长短等,这些都将决定所研究的是哪一组具有相似表现的患者。本例中可选择性别不限、年龄在 18~70 岁活动

期类风湿关节炎患者作为研究对象。

I(intervention): 干预措施,需评价的措施,如本例中的新型口服药物。

C(comparison): 参照措施,通常选择目前指南高级别推荐的或得到专家共识的标准治疗措施,如果确实找不到有证据的治疗方法,在不违背医学伦理学原则的前提下,可选择安慰剂作为对照组。目前甲氨蝶呤是治疗类风湿关节炎的锚定药物,应用甲氨蝶呤治疗是较理想的比较措施。

O(outcome): 结局,疾病治愈率、缓解率、生存率、死亡率等终点指标均可以作为主要结局,建议采用有很好反应性、稳定、价廉、定量和经济的指标评价。如对于疼痛等主观指标的评价可以采用定量或半定量评价方法,如视觉模拟评分法(VAS)、口述分级评分法(VRSs,无痛、轻微疼痛、中等度疼痛、剧烈的疼痛,每级1分)。本例根据具体临床问题不同,即可以选择疾病活动指数下降等药物有效性指标(DAS28、SDAI、CDAI等)作为结局,也可以选择不良事件发生率等药物安全性指标作为结局。

PICO模式使研究者能确立核心的问题和问题类型,这些信息有助于寻找最适合的证据或研究类型。如能为药物疗效提供高级别证据的研究方法为基于随机对照试验(randomized controlled trial,RCT)的Meta分析和系统综述或者大样本量的RCT,但以药物安全性为主要结局的研究常选择观察性研究方法如病例对照研究(case control study)。

第二节 治疗性研究方法的选择

按照研究设计方法进行分类,治疗性研究可以分为观察性研究和实验性研究两大类。

观察性研究包括病例报告或病例系列观察报告、队列研究、病例对照研究等。观察性研究具有易于开展、研究时各方面受到的限制少、所耗资源亦相对较少等优势,但观察性研究不可避免地会受到各种非主要研究因素或非研究因素的影响,因而研究设计和实施应该尽可能周全、严谨,对其结果的解读亦应格外审慎。

实验性研究包括随机和非随机临床试验,其中随机对照试验在各种临床疗效考核的方法中具有最高的论证强度,能最真实地反映所研究治疗措施的临床疗效。其缺点是在具体实施时有一定的难度,对伦理学的要求更高。

对于某种治疗措施最早的研究往往是观察性研究,通过早期的调查和观察,往往能形成对这种治疗的初步印象,并产生科学的假设(hypothesis)。研究者根据提出的问题,设计开展与背景知识相对应的各种不同类型、具有不同特点的实验性研究,比如新药研发和应用中的Ⅰ~Ⅳ期临床试验,临床试验的主要研究设计为RCT。Ⅰ期临床试验是初步的临床药理学及人体安全性评价研究,观察人体对新药的耐受性和药代动力学,为制定给药方案提供依据。Ⅱ期临床试验是治疗作用的初步评价阶段,目的是初步评价药物对目标适应证患者的治疗作用和安全性,也为Ⅲ期临床试验设计和给药剂量提供依据。Ⅲ期临床试验是治疗作用的确证阶段,目的是进一步验证药物或某种干预措施对目标适应证患者的治疗作用和安全性,评价利益与风险关系。Ⅳ期临床试验是上市后研究,目的是考察在广泛使用条件下的药物疗效和副反应、评价在普通或者特殊人群中使用的利益与风险

关系以及改进给药剂量等。

因此,治疗性研究方法的选择是需要掌握不同研究方法的优缺点,考虑当前对所要研究的临床问题认识的深度和广度,掌握相关领域背景知识的研究进展。

第三节 治疗性研究科学性评价

一、随机化

通过临床随机分组,每个患者有同等的机会进入治疗组或对照组。其目的是治疗组和对照组中除研究的治疗因素外,其临床特征、预后和其他因素的分布应该均衡,有可比性,这样有可能得到真实的研究结果。在非随机试验中,医师有权决定患者进入治疗组或对照组。有预后较好的患者常常会被分配到治疗组中去,这样得到的治疗性疗效往往比随机分配所得到的结果更有效,有时甚至会有假阳性的结果。因此,随机化是评估治疗临床试验时第1条而且是最重要的标准。

1. 随机化分组

(1) 简单随机化:根据计算机所产生的随机数字或随机数字表获得。当样本量足够大时如总例数大于200例的治疗研究,发生组间病例数分配显著不平衡的机遇可以忽略不计。但样本量较小时,应用该方法进行随机化有一定概率产生组间分配不平衡达到3:7,甚至更大。

(2) 随机区组法:根据研究对象进入试验时间顺序,将全部病例分为含量相同的若干区组,每一区组内病例随机分配到各研究组,这样可以避免两组之间人数的差异过大。

(3) 分层随机化:为了减少重要的预后因素可能在两组分布不均匀,可以根据预后因素分层,在每层内将患者随机分配到治疗组和对照组。如类风湿关节炎治疗后放射学进展评价,可以以抗CCP抗体高滴度和低滴度分层,病程小于2年和大于2年分层。需注意,分层越多,划分的区组也越多,分层过多常常会导致每一层的治疗组和对照组人数过少,不利于统计分析。

2. 隐藏随机分组(concealed random allocation) 是保持随机化过程完整性和保证预后因素在两组均衡相等的一项措施。通常情况下,制定随机分配方案的研究者不参与确定受试者是否适合入选,也不参与实施治疗方案或者分析结果。在实施隐藏随机分组的研究中,确定受试者入选的医师不知道进入试验的下一个患者会被分配到治疗组还是对照组。可以避免:①患者主观要求进入某组;②医师的主观愿望。从而可以消除选择偏倚。没有隐藏的随机分组,医师可能会根据患者情况决定其进入哪一组,随机化遭到破坏。已有研究表明没有隐藏的随机对照研究得到的某药的治疗作用大于使用隐藏随机分配的试验所得出的结果。实现充分的随机分组隐藏的措施还包括顺序编号、不透明、密封的信封,顺序编号的容器,药房控制和中心随机化等。

在评价临床随机对照试验的文章时,应该明确研究者使用上述哪一种随机化方法和隐藏随机分组的措施。如为多中心药物临床试验,应阐述:①中心如何进行药物编号;②治疗组药和对照组药的编号装入信封密封;③如何完成密封的信交给各分中心研究

机构;④ 各分中心如何将药物发给入选对象;⑤ 遇到什么情况下需要揭盲。

二、盲法

盲法(blindness)指患者、医师或参加试验的研究者(包括实验室人员、统计学家)不知道患者接受的是治疗药还是对照药。盲法可以消除患者和医师的测量偏倚和霍桑效应,特别对于判断的结果是主观评价的指标时(如关节痛、头痛、乏力)尤为重要。但当双盲不可能实行时,例如评估类风湿关节炎关节放射学进展,这种情况下可将影像学片子集中在一起,让一组不知道这些患者采用什么措施治疗的专科医师进行盲法评定,得到的结果将更可靠。

根据设盲对象不同分为:① 单盲试验,受试者或研究人员一方不知道所接受治疗;② 双盲试验,受试者和参加临床试验或临床评价的研究人员或申办者方工作人员均不知道也不能识别研究对象接受了何种治疗。在试验实施过程中一直保持盲态,只有在试验结束、完成数据清理、数据已达到可以接受水平并且“锁定”后方可由指定人员揭盲。

当出现比较的两种药物剂型不同并可能会引起药代动力学和(或)药效动力学的改变、两种制剂的给药途径不同(注射剂和口服剂型)等情况,要达到理想的双盲会遇到困难。这时,可采用“双模拟”(double-dummy)技术,即为每一种治疗方法准备一个相同剂型、相同给药途径的安慰剂,如在新型口服片剂药物治疗类风湿关节炎的双盲双模拟临床试验中,治疗组接受新药药物片和甲氨蝶呤模拟片,对照组则给予新药模拟片和甲氨蝶呤药物片。

三、设立对照

对照组(control)是除治疗干预措施外,其他条件与治疗组一致的一组研究对象,在RCT中,符合入选标准的患者随机进入治疗组或对照组,使两组患者临床特征均衡可比,消除非干预因素对患者结局的影响,有效评价试验措施的真实效果。

1. 安慰剂对照 在药物疗效评价中,常常给对照组服用在外观、气味与考核药物相同但没有药理作用的安慰剂。从伦理角度来讲,这种情况仅用于目前缺乏有效治疗措施的临床治疗性研究或新药Ⅰ期临床试验。

2. 自身对照 同一组患者先后接受两种不同的治疗,以其中一种治疗作为对照,比较两种治疗结果的差别,以确定所考核药物的疗效。这种研究设计适用于慢性稳定或复发性疾病,如轻中度高血压和高血脂等。由于同一组病例先后作为治疗组和对照组而接受治疗,可确切判断每例患者对研究因素和安慰剂的反映,故具有良好的可比性,结果的可靠性亦远高于不同病例组的前后对照研究。缺点是每一例的研究期限延长一倍,患者的依从性容易受到影响。在前后两个治疗阶段之间,有洗脱期(washout period),时间长短依所选药物的半衰期和病种、病情而定,然后开始第二阶段治疗。尽管如此,第一阶段药物作用有可能影响第二阶段研究;如第一阶段已治愈或一般情况发生改变或死亡的病例则不能进入第二阶段。

3. 交叉对照(crossover) 是对两组受试者使用两种不同的治疗措施,然后相互交换处理措施,最后将结果进行比较的试验方法。优点是每例患者先后接受治疗组或对照组的的治疗,消除了不同个体间的差异。随机分组可避免组间差异和人为选择偏倚,且需要

的病例数较少。缺点是应用病种范围受限,对于各种急性重症疾患(急性胰腺炎等)或不能回复到第一阶段治疗前状况的疾病(如系统性红斑狼疮)以及无法停止治疗让病情回到第一阶段的疾病(如心力衰竭)等,都不能采用交叉对照试验;该研究同样需要洗脱期,由于两个阶段的治疗可能有重叠,故需要一个洗脱期;整个研究观察期较长,患者的依从性不容易得到保证。

4. 历史性对照(historical control) 将现时给予所研究药物治疗的一组患者的结果与既往治疗的一组患同种疾病但未给予该药治疗的(安慰剂或传统药物)患者的结果相比较,以评价该药的疗效。缺点是特别容易产生偏倚,不能保证今昔两组患者的病情和所考核的药物以外的治疗具有可比性。

四、除干预措施外,两组治疗的一致性

当治疗组对象额外地接受了有利的治疗,结果夸大了该治疗措施的有效性,称为干扰(co-intervention)。例如在进行的痛风患者有关降尿酸药物治疗疗效评价中,治疗组获得更多的一般治疗,包括饮食治疗,避免诱发因素或药物等,影响了疗效的真实性。

如果对照组额外地接受了治疗组措施或其他有利的治疗,人为地夸大了对照组的疗效,称为沾染(contamination)。例如纳入类风湿关节炎患者在不同医院或不同医师处诊治,接受了额外的改变病情药物,缩小了评价药物和对照药物疗效的差距。

五、随访的完整性

被研究患者随访是否完整,对于决定结果评定的可靠性是十分重要的。随访的完整性需要考虑根据不同疾病要有足够的长度,接受药物起效时间,副反应谱出现时间点,如类风湿关节炎患者接受甲氨蝶呤治疗,起效时间通常在治疗后3个月,随访时间不应小于3个月,药物副反应如胃肠道不适、脱发、血细胞减少等,出现较早,而对肝脏毒性作用与治疗累积剂量有关,一旦累积剂量达1500mg,需要严密监测。因此,针对甲氨蝶呤药物作用机制、代谢特性,对该药考核时间不能少于3个月。

失访的含义是在试验某一时间点上,需要测定患者的结果时,却无法联系到患者。对于这种情况可能因素是:①治疗有副反应,患者不愿继续接受治疗,或者患者在随访这段时间中已经死亡;②由于症状已缓解,患者不愿继续治疗或随访;③患者由于搬迁离开原地址;④拒绝接受某些检查,特别是创伤性检查,例如非甾体抗炎药对胃肠道损害需要采用胃镜检查等。一般而言,如失访率超过20%,则研究结论真实性会受到质疑。研究者应该努力提高随访率,在随机分组前剔除可能不愿参加随访的受试者,获得受试者的联系方式,随访时制定流程以利于快速完成每一次随访,随访时间尽可能简短,提供优质免费的医疗服务,提供资金资助如患者交通费用等。

六、评价结果是否采用意愿治疗分析方法

在评定治疗性文章的科学性时,应检查该文章是否采用了意愿治疗(intention-to-treat analysis, ITT)分析方法。所谓ITT分析是指所有纳入随机分配的患者,不管他是否最终接受分配给他的治疗,在最后资料分析中都应被包括进去。ITT分析可以防止预后较差的患者,在最后分析中被排除出去,可以保留随机化分配的优点,即两组可比性,使结论更可靠。

而按方案(per protocol, PP)分析是通过将研究对象限定于那些完全遵循医嘱的对象来确定进入最终资料分析的。因此在 PP 分析时需剔除失访者的资料,计算时的人数仅为随访完整的患者。

两者不同的是计算疗效时的分析在 ITT 时为所有的人选人数,而 PP 仅为剔除失访以后的人数。由于 PP 分析时剔除了不依从试验而没有完成试验的人,而不依从者常由于药物副反应或治疗后疗效差而离开试验,因此,PP 可能会过高地估计治疗结果。失访者越少,PP 与 ITT 的结果越接近。在判定失访对于结果评定的影响时,如果治疗组优于对照组,进一步将治疗组失访的人都算作无效,而对对照组失访者都算作有效,重新计算,若重新计算后,结果仍旧是治疗组疗效优于对照组,则说明失访并未对最后的结论产生影响。

七、治疗性研究疗效大小的评估

在评估研究结果即疗效大小时,应考虑到其在临床意义和统计学上的显著意义。有时虽然在统计学上有显著意义,但结合临床分析并无显著的临床意义。

1. 危险度(risk) 估计治疗效果的指标,某一个评价结果的频数分布,波动于 0.0~1.0 之间。若为 0.0 表示事件不会发生,为 1.0 表示事件必然会发生。对于临床疗效大小的估价参数有相对危险度(relative risk, RR)、相对危险度减少(relative risk reduction, RRR)、绝对危险度的减少(absolute risk reduction ARR 或 RD)。

(1) 相对危险度:即治疗组相对于对照组的危险度,是两组危险度之比,也就是治疗组发生不良事件(adverse outcome)的危险度除以对照组的危险度。 $RR < 1$ 说明治疗组的干预措施能降低不良事件发生的危险度; $RR > 1$,说明干预措施反而增加不良事件发生的危险度,因此 $RR > 1$ 说明干预措施是有害的。

(2) 相对危险度的减少(RRR): RRR 表示了治疗组与对照组相比,其不良事件减少的相对数。即治疗组比对照组减低的危险度除以对照组危险度, $RRR = (\text{对照组出血率} - \text{治疗组出血率}) / \text{对照组出血率}$ 。

(3) 绝对危险度减少(ARR): ARR(或 RD)是治疗组和对照组不良结果事件的绝对差别。ARR 为对照组事件发生率(危险度)减去治疗组事件发生率(危险度),在本例中 $ARR = 61\% - 26\% = 35\%$ 。由于 RRR 仅是相对数,有时危险度下降的绝对数很少,RRR 看上去却很大,会引起误导。

2. 需要治疗的患者数(need to treatment, NNT) NNT 是 ARR 的倒数, $NNT = 1/ARR$ 。意思是为了预防 1 例不良结果事件需要治疗的患者数。例如,我国多中心有关比较来氟米特(LEF)和甲氨蝶呤(MTX)治疗类风湿关节炎的临床试验,在 12 周时 NNT 为 11.9,即 LEF 每治疗 12 例患者就较 MTX 新增加 1 例治疗有效病例。NNT 适用于评价治疗病情相同,并且取得相同结果的各种治疗方法。可以根据 NNT 对这些疗法进行等级评定,这将有利于提高治疗效果的评估,当然由此得到的 NNT 等级并不一定直接决定临床决策,因为还必须考虑到药物的副反应、价格、患者的性格、期望和选择,更重要的是,有时不接受治疗患者也会有有效结果发生,而且这种结果发生的频率会影响 NNT。

3. 最小临床重要差异值(minimal clinically important difference, MCID) 是评价

药物治疗疗效是否具有临床意义的定量指标,常用的确定某种疾病 MCID 的研究方法有效标法、分布法、专家意见法和文献分析法。不同的疾病和研究终点有不同的 MCID,有时标准治疗的改进也会导致 MCID 发生变化。当一项治疗研究获得有统计学意义的结果后,还应该分析治疗组与对照组的组间差异是否达到了 MCID,为筛选治疗药物、再评估样本量和解释试验结果提供证据。类似的指标还包括最小临床重要变化值(minimal clinically important change, MCIC)和最小临床重要改善值(minimal clinically important improvement, MCII)。MCIC 与 MCID 的区别在于 MCIC 评估的是单个受试者治疗前后发生的变化,在治疗性研究中,研究者进一步比较治疗组与对照组达到 MCIC 患者的百分率的差异,从而判断试验药物的临床意义。

第四节 如何将治疗性研究结果用于患者

一、将研究结果用于自己的患者

应全面评价患者的情况,包括生物学指标、重要预后因素、信仰、治疗倾向、经济水平等。如果一项研究的患者入选和排除标准与自己的患者相同,那么该项研究结果应用于自己的患者有可能获得相应的效果。如果入选标准与自己的患者略有不同,例如有并发症、年龄较大或病程更久,将该方法应用自己的患者需要谨慎。

有时对于进入研究的所有患者来讲并不能发现治疗措施有效,但是对其中某一亚组却是有效的。在评价亚组疗效时,尚需了解亚组病例数如何,副反应有否增加。

二、这项治疗能否在自己的患者中应用

文献确定该项措施的确有效后,需要进一步评价该措施能否在自己患者中应用,这涉及患者及医疗保险系统能否支付该项治疗措施的费用,所在的医疗单位有无开展该项技术,有无配备治疗药品?是否开展相关检查进行监察和随访?如不可能时就有可能不能将该治疗措施用于自己的临床患者。

三、治疗措施对患者的利与弊

在对患者进行一种新的治疗时,医师和患者都会考虑到这种治疗方法对患者的好处与坏处,在考虑对患者的有益作用时,必须考虑到临床上所有重要的结果。

在考虑治疗措施利和弊时,NNT 和 NNH 是较好的衡量指标。对于同一种疾病、同种结果的不同干预措施,NNT 有助于对治疗措施作出临床决策,NNT 愈大,治疗效果愈好。当然在具体落实到每个患者时,应全面考虑各方面因素。如果治疗或预防措施费用较低,安全性大,使用方便,加上尚无有效的措施或如果不用该项措施又可能出现严重后果,这种情况下,较小的 NNT 也可以接受。在实施某项措施前,必须将治疗的必要性和治疗效果给患者交代清楚,还要交代不进行此项治疗,可能会发生什么不良后果。再告知进行这项治疗可能发生的意外和副反应以及费用,让患者自己作出决定。

第五节 在实施有关治疗措施的循证医学时的参考步骤

- (1) 应用 PICO 模式提出临床问题。
- (2) 根据临床问题找出最恰当的相关的研究文章。
- (3) 评价治疗试验文章的科学性
 - 1) 是否进行随机化分组?
 - 2) 是否实行隐藏随机分组?
 - 3) 是否对患者和研究人员双盲?
 - 4) 对照组的设立是否合理?
 - 5) 除了需要评估的治疗措施外,两组是否得到相同的治疗?
 - 6) 被研究患者的随访是否完整?
 - 7) 资料的总结是否采用意愿治疗分析?
 - 8) 临床试验设计、实施和评估过程中是否符合医学伦理学原则?
- (3) 临床上和统计学上有显著意义的结果
 - 1) 估计治疗效果的大小。
 - 2) 样本大小的评价。
 - 3) 治疗作用的精确性。
- (4) 将临床结果用于自己的患者
 - 1) 研究结果是否可用于自己的患者?
 - 2) 这种治疗方法可否应用?
 - 3) 治疗措施的好处与坏处及费用相比是否值得应用?

(姜林娣)

【思考题】

1. 试述随机化的意义。
2. 用于评估治疗性研究作用的大小结果指标有哪些?
3. 实施有关治疗措施的循证医学步骤包括哪些?