·系列讲座·

临床流行病学

第三讲 临床流行病学研究中的因果推论

李 辉

在验证病因及防治效果的评价中，涉及识别和论证因果关系的问题，即回答不同事物间的联系究竟是表面上的联系还是本质上的联系。前者是一种虚假的联系，后者是一种因果上的联系。虚假的联系是应极力避免的，应尽力发现各种虚假的联系，以免产生错误的因果结论。因果联系是期望发现和证实的客观规律，是研究的根本目的。临床流行病学研究中的因果推论是指从实验资料中产生正确结论的过程，不仅要从错综复杂的事物关系中，排除各种偏倚和误差的影响，而且还要经过严格的逻辑检查，多方论证才能下结论。确定因果关系的推论过程基本步骤如下。

一、确定两事物间是否存在统计学上的联系。在现场实验（或临床试验）中，评价干预（或预防）措施及药物（或疗法）治疗的效果也是经常遇到的研究课题，反映干预效果及疗效好坏的指标如率差 RD 值，主要是由三方面的因素决定：①干预措施（或疗法）的真实效果；②偏倚效应（影响干预或治疗效果的因素在实验组和对照组中不可比所致）；③随机误差。样本 RD 值的统计检验是在没有偏倚效应的影响下，确定该 RD 值由随机误差所致的概率（即 P 值），以此对干预措施的效果做出评价。

鉴于上述道理，对所获样本结果进行显著性检验，无论在统计学上是否有显著意义（$P\leqslant0.05$或 $P>0.05$），均不能盲目的仅根据统计学上有无联系便下因果结论，首先应对此结果是否为各种偏倚所致。或部分受偏倚的影响做出判定。

二、确定偏倚对所获研究结果的影响。当样本结果有统计学上联系时，不应盲目乐观。因所获的 RD 值有可能是受偏倚（选择偏倚、测量偏倚或混杂偏倚）影响的结果，并非真实的因果关系。因此，遇到 $P\leqslant0.05$时，一件非常重要的工作便是认真寻找该项研究的全过程中（从设计、测量直到结果分析），是否存在偏倚的影响。如确实存在偏倚，且影响较大时，应详细分析所获资料受偏倚影响的原因。若存在严重的选择性偏倚，结果很难再校正，一般应重新设计再次做此项研究，或如实报告此结果，并指出其受选择偏倚影响的程度，同时不宜将此结果外推到总体人群。若存在测量偏倚，并知道测量误差的方向和大小（如病例诊断定义或疗效评价指标测量方法的敏感性和特异性）时，可对观察结果进行定量调整；若预备试验中未对该项研究中所采用方法的信度和效度做过估计，则调整无法进行。此外，有些指标若有时点性时，则无法通过重复测量来校正不准的结果。若存在混杂偏倚，一般可采用统计方法加以调整，去除混杂效应的影响，但必须对混杂因子在将要比较的两组（实验组和对照组）中的分布情况有客观的如实的测量，否则无法调整混杂因子产生偏倚的影响。如果经上述检查后，所获结果确实不受偏倚影响，则研究结果有可能存在因果关系的可能性，可做进一步的因果论证。

当样本的结果经统计学检验无显著意义（$P>0.05$）时，同样应按上述原则和步骤进行分析，不要盲目悲观。因上述偏倚可以使一原本存在因果关系的客观事实，在统计学检验时没有统计学显著意义。如找不出偏倚的影响，则还应对此结果是否受样本数量过小的影响加以分析，必要时可再增加样本数量进行观察。如扩大样本后，所获结果仍表现不出统计学上的联系，则可对此研究下结论，认为“因”与“果”之间不存在因果关系。

三、按判断因果关系的原则确定研究结果的因果性质。经临床试验和现场实验所获的任何一项研究结果，若要对结果下因果关系的结论，除了要保证此结果不存在任何偏倚影响和统计检验表明存在显著统计学意义之外，还应按严格的判断因果关系的

作者单位：中国协和医科大学中国医学科学院基础医学研究所 北京 100005

一些原则进行逻辑关系的检验。如果所获研究结果与某一判别因果关系的原则不相符合，则仍然不能下因果结论。若结果与这些检验的原则相一致，此时才能较有把握地对此问题的研究结果下因果结论。

临床试验和现场实验方法主要用于评价疗效、验证病因假设或评价预防措施的效果。验证病因假设，可通过阻断已暴露于某一危险因素的人群再继续暴露于该危险因素，最后观察该人群产生所研究疾病的频率，是否比对照人群低。如 Doll 和 Hill 在研究肺癌与吸烟的关系便运用了这种人群实验。而评价预防措施效果，则是通过对易患某病的高危人群，人为地施加相应预防措施，最后观察此人群产生该病的频率是否比对照人群低。各种人群疫苗接种效果评价，均是运用了这种人群实验的方法。阻断继续暴露于危险因素或施加保护因素，统称干预措施，是研究者人为地施加于实验人群的。干预措施不同于观察性研究中的暴露因素，后者是自然状态下暴露于人群。因此，临床流行病学实验研究可按严格实验设计进行实施，干预措施的实施与结局的发生并不存在时间顺序颠倒的矛盾，因而措施与结局之间联系的因果关系似乎一目了然。但结论仍易受诸多因素影响，如疾病病因多不是单一的，各危险因素（或致病因子）间常存在竞争风险（competing risk）或交互作用（interaction effect）；许多疾病的自然史尚不完全清楚；特别是干预实验的组织实施往往非常困难，定量而准确地测量干预强度尤其艰难。至使在对现场实验的结果下因果结论时，除需对所获资料的完整性和真实性进行严格检查外，还应对要产生的结论进行严密的逻辑检验。有关检验的原则与内容如下。

（一）样本的代表性。由于实验是人为地将干预措施加于实验人群，因而受试者的选择往往难于按随机抽样的方法获得。即便样本有较好的随机性，有时也常因受试对象拒绝接受干预措施或实验过程中退出实验，导致样本中的受试对象过多丢失；有时为保证研究的可行性，不得不放弃随机性而选择部分易于合作的受试者。固从具有上述情况的样本产生结论时，须注意样本代表性是在什么水平上，特别是在下因果结论时尤其要慎重。一个来自有代表性的样本的结论，往往和别的研究者在不同的时间、不同的地区和人群，进行的同类水平的研究，有很好的一致性。产生因果结论前，应描述和检查样本的代表性，具体内容如下：

1. 如果抽样过程中，某些抽样单位（受试者）无法按抽样方案规定的随机原则抽取，应描述这类为满足研究可行性而放弃随机性所选择的受试者数占样本容量的百分比。如太高则该样本所产生的研究结论未必同于总体结论。

2. 如果实验进行过程中，有相当比例的受试对象退出实验或失访，应确定在实验终点时还剩的受试对象，在数量上是否满足实验设计所估计的最低样本容量的要求。若样本丢失太多，从样本所获的结果未见统计上的显著差别，不应下阴性结论和外推到总体人群。

3. 如果所抽取的样本与总体相比，在某些人口学指标方面（如年龄、性别及职业构成、出生率、死亡率及人均收入等）存在较大差别，样本在对干预措施的反应方面有可能不同于总体人群，从样本获得的结论外推总体人群时应慎重。此外这类指标可能不全代表受试者对干预措施的敏感性方面的情况，因而这类间接指标有时根本不能用来评价样本是否能代表总体。

4. 如果受试对象不是按完全随机的原则来确定，而是按照某种人为定义来选择，多存在选择偏倚，样本因而无代表性可言。利用自愿者进行研究在下总体人群因果结论时应慎重。

（二）实验组与对照组的可比性。没有比较就没有鉴别。实验是通过比较实验和对照两组的结局指标来产生结论。然而，比较只能在相同条件下比较。这里所说的相同条件指的是除了施加的干预措施（或治疗药物）之外，其它一切对干预的结局有影响的因素，在实验人群和对照人群中均应一样。所谓的两组间的可比性指的就是这些其它的因素，在两组之间的均衡程度。缺乏可比性的实验结果可能存在不同程度的混杂偏倚，往往会导致错误的结论。在对干预措施和干预的结局之间下因果结论时，必须要对两组进行可比性检验，检验的内容如下：

1. 如果要研究的问题是验证病因，而研究的疾病又可能是多病因的疾病，必须注意竞争风险对结果的影响。现以通过戒烟人群实验来验证吸烟与肺癌因果关系的研究为例，说明其重要性。在这一研究中，戒烟的实验人群和不戒烟的对照人群均有可能受其它的肺癌危险因素和保护性因素的影响。如这两人群暴露于这些危险因素或保护性因素不可比，则会使研究结果出现多种情况的偏倚。现用公式来描述戒烟干预的效果如下：$D=R_I-r_c$ 式中：R_I：实验（戒烟）人群的肺癌发病率；r_c：对照（不戒烟）人群的肺癌发病率；D：实验人群与对照人群肺癌发

病率的差值。若戒烟可减少肺癌的发生，则D值为负值（D＜0）。现将R_1和r_c的归因做一分析。R_1可认为由下述几种因素所致：$R_1=R_1+R_2+R_3+R_4$ 式中R_1：代表戒烟人群戒烟前吸烟的致肺癌作用；R_2：代表其它已知的各种肺癌危险因素的致肺癌作用（增加发病，为正值）；R_3：代表已知的各种保护性因素减少肺癌发生的作用（减少发病，为负值）；R_4：代表未知的肺癌危险因素和保护性因素的总效应。r_c由以下几种因素所致：$r_c=r_1+r_2+r_3+r_4$ 式中：r_1：代表对照人群吸烟的致肺癌作用，包括戒烟实验开始前后全部吸烟致癌作用的总和；r_2：代表其它已知的各种危险因素致肺癌的作用（增加发病，为正值）；r_3：代表各种未知的肺癌保护性因素减少肺癌的作用（减少发病，为负值）；r_4：代表各种未知的肺癌危险因素和保护性因素的总效应。

由于在实际工作中无法确定真实的R_4及r_4值，因此假定各种未知危险因素和保护性因素的效应是相等的，以便于对D值讨论简单化。下面的讨论是建立在$R_4=r_4$基础上的分析。

本项研究的目的是期望戒烟能使肺癌发病率降低，从而推论吸烟是肺癌的病因。也即评价的指标D＜0。但从上述分析可看出，D值受R_1-r_1，R_1-r_2及R_3-r_3差值的影响。要使D值只反映戒烟的效果，R_2-r_2及R_3-r_3均必须等于0，即危险因素和保护性因素在两人群中产生的效应均应相同，否则将影响D值的真实性。在实际工作中，除吸烟之外其它导致肺癌发生的危险及保护因素影响情况可作如下推论：

①假定$R_3-r_3=0$，D值不受保护性因素影响时，如$R_2>r_2$（R_2及r_2均表示危险因素产生的发病效应），即戒烟人群因该因素产生肺癌人数比对照人群多，表现为$R_2-r_2>0$，D值增大，戒烟产生的减少肺癌发病的真实效应被缩小。

②假定①中条件不变，如$R_2<r_2$，则$R_2-r_2<0$，即D值被缩小，戒烟的真实效果被扩大。

③假定$R_2-r_2=0$，D值不受危险因素影响时，如$R_3<r_3$（R_3及r_3表示的均是保护性因素产生的防病效应），即戒烟人群因该因素减少的肺癌人数比对照人数多，表现为$R_3-r_3<0$，D值缩小，戒烟产生的减少肺癌的真实效应被扩大。

④假定③中条件不变，如$R_3>r_3$时，则$R_3-r_3>0$，即D值扩大，戒烟真实效果被缩小。

⑤若$R_2-r_2\neq0$且$R_3-r_3\neq0$时，还有三种情况对D值产生影响，这取决于$R_2-r_2=R_3-r_3$，还是$R_2-r_2>R_3-r_3$或$R_2-r_2<R_3-r_3$，这里不再做详细讨论。

上述各种情况中R_2、r_2、R_3、r_3的效应在实际观察中很难与R_1、r_1区分，完全建立在两组中的R_2及r_2值与相应危险因素的暴露频率相关，R_3及r_3值与相应保护性因素的暴露频率相关基础上，分析时仅仅是根据危险因素和保护性因素在两组中暴露的频率来间接估计它们的效应。干预试验强调的实验组与对照组间均衡性，实际上是这些影响结局因素在两组分布频率上无差异，而未必代表诸因素真实效应（影响结局的效应）在两组相同。根据这类实验结果下结论时同样应留有余地，因这些因素对干预所产生的影响毕竟缺乏真实效应的测量证据来证实。

从上述分析还可看出，如果吸烟者尚未因吸烟而致肺癌，而是由其它致肺癌的危险因素抢先一步产生了肺癌（假定该肺癌患者对其它致肺癌危险因素更敏感；假定吸烟与其它危险因素间无交互作用），这种效应被称之为竞争风险性。从广义来讲，凡是可使吸烟产生肺癌的效应不能及时表现出来的一切因素，均可具有竞争风险性。如一个吸烟者原本会产生肺癌，但因突然患心肌梗死或车祸而提早死亡，使吸烟将导致的肺癌无法再表现出来。在实际研究中，肺癌其它危险因素所致的竞争风险性对研究结果的影响，往往是很难准确测量和估计的，只能通过比较实验人群和对照人群暴露于此因素的频率而进行间接地估计。按上述竞争风险的定义，与吸烟致肺癌相竞争的因素可分三类：ⓐ导致其它死因的因素；ⓑ其它致肺癌因素；ⓒ肺癌保护性因素。

综上所述，根据实验的观察结果产生因果结论时，即便保证了某些危险因素和保护性因素在两人群间分布频率的可比性，仍然要非常慎重。原因一是某些未知因素的竞争风险性在两人群中是否相同，目前进行的研究无法提供任何信息，即便实验人群和对照人群是按随机分配的原则来确定的，也存在缺乏可比性的某些可能性；二是两人群间危险因素和保护性因素的均衡性测定，只能间接估计竞争风险性在两人群间是否可比；三是某些导致在观察终点前丢失样本的因素，其产生的竞争风险性，在两人群间的均衡性很难人为地控制。特别是对于肺癌这样一种小概率事件作为评价的结局指标时，可能会产生较大的偏倚。

2. 如果要研究的问题是预防接种或药物预防的效果评价，则应注意暴露于致病因子的机会和敏感个体比重大小对结果的影响。无论是接受疫苗或药物的实验人群还是对照人群，所要预防疾病在两人

群的发生均需具备下述三个条件：

①致病因子：要预防的疾病如是传染病，则致病因子应是病原微生物，如是非传染病则为各种无生命活动的环境因素（包括社会因素）及人的精神、心理或行为因素。若实验的地区或人群（实验和对照人群）中根本不存在该病的致病因子，则会导致阴性结果，即预防接种或药物干预无效。选择研究人群时应予注意。

②暴露于致病因子的机会：若实验地区或人群中存在该病的致病因子，但是实验人群和对照人群暴露于该致病因子的机会不同，则会导致两人群因暴露频率不同而发病率不同（假定暴露频率与发病率成正比）。实际工作中，用何种客观的指标来描述实验人群和对照人群暴露机会的可比性，应根据具体的研究内容来确定。如评价麻疹疫苗效果，可以使用历史上两地区两人群麻疹发病率或其它呼吸道疾病的发病率（如百日咳、猩红热等），也可利用实验开始后，通过前瞻性观察，确定两人群急性呼吸道感染（如急性咽炎、感冒等）的发病率，以这类指标来反映两人群的流动、交往接触情况，间接地反映两人群暴露于麻疹病毒的机会是否可比。如评价口服常规剂量的叶酸是否能有效预防神经管畸型的发生，可通过测量两地区两人群日常饮食中缺乏叶酸的孕妇人数及其缺乏叶酸的程度，以此指标来反映两人群的暴露机会是否可比。此外，保证实验人群和对照人群暴露机会可比性的办法可在设计阶段加以控制。一般来说按地区的整群随机化分配确定的对照，属于外对照的形式。在保证人群暴露机会可比性方面，其效果远不如采用内对照的形式好。内对照指的是将同一地区、同一单位（如学校班级、工厂车间、机关办公室等），甚或居住在一起的家庭的成员，按随机化分配的原则，确定为实验人群或对照人群。

③敏感个体：所谓敏感个体指的是对致病因子没有先天及后天抵抗能力，一旦暴露于致病因子便会发生相应疾病的个体。如实验人群和对照人群虽暴露于致病因子（研究期望要控制的病因）的机会相同，但两人群中敏感个体的分布不同，仍会影响实验结果的真实性。在实际工作中，如何确定敏感个体，是一项比较复杂的工作。对于疫苗评价来说，评价的指标可用血清抗体水平，但所有观察人群均测量无疑导致高花费问题，解决办法可在两人群中各抽取一样本来估计敏感者的比例有无差别。对于非传染病来说，确定敏感个体是一项尚待探索研究的新领域，分子生物学技术定将提供有价值的指标。目前通常使用严格的随机化分配来保证敏感个体在两人群中的分布达到均衡。有些指标虽然被提出来作为预测个体是否易于产生所研究的疾病，如盐敏感个体有可能易于产生高血压，但有两个难点尚未解决。一是高血压病至今被学术界认为是一种多病因的疾病，高盐只是诸多危险因素之一，对盐敏感是否也能代表对其它危险因素敏感尚无定论。其二是高盐摄入与高血压病的关系，仅停留在统计关联这一水平上，即高盐饮食只是高血压病的危险因素而已，离生物学上下因果结论尚缺乏更多的证据。总之，在非传染病的干预实验中，如何确定敏感个体尚无很好的指标来测量，因而有关非传染病干预实验的结果，在产生因果结论时应非常慎重。其远不如疫苗评价，能保证排除因敏感个体在两人群分布不均所致的偏倚。

综上所述，根据人群干预实验产生因果结论时，保证实验人群和对照人群在疾病发生和流行的三个基本条件：致病因子、暴露于致病因子的机会和敏感个体的均衡性，是评价两人群可比性的另一重要内容，是评价实验结果能否产生因果结论的重要依据。

3. 人群现场实验无论是用于验证病因还是用于评价疫苗或药物等某一干预措施的效果，均涉及结局指标（如患病）的定义及测量方法的可比性问题。例如病例的诊断标准、诊断方法在两人群是否一致、是否统一，以及从研究开始到研究结束，是否均按标准化的诊断标准和方法去执行等，与前述原则一样重要。若结局指标的定义、测量方法没有标准化，或在整个研究过程中没有实施质量控制，同样都会使观测结果失去真实性。特别是当两人群在标准化和质量控制问题上缺乏可比性，会使结果严重失去真实性。在实际工作中，克服这种人为造成的测量偏倚是较容易实现的。对设有对照的比较性现场干预实验来说，诊断标准和诊断方法在实验人群和对照人群间缺乏可比性，要比诊断标准比较粗糙，诊断方法敏感性和特异性不太高（但两人群间完全可比）产生的偏倚可能更大。因此，从实验结果产生因果结论时，在评价的结局指标定义和测量方法方面，也必须进行可比性的检验，以保证结论不受测量偏倚的影响。此外，如结局指标不是严重的临床情况时，较轻的临床症状和体征往往容易造成受试对象或观察员的忽视，如两人群在发现或报告病例方面缺乏可比性，同样会产生上述偏倚。因而在确定因果关系时，这种指标也应列入检验两人群可比性的依据。

（三）依从性。依从性指的是实验组接受干预措

施的落实情况，广义来讲还应包括对照组成员是否也接受干预措施的情况。实验组成员部分未接受干预措施（或接受干预措施的强度不够）可比喻为实验组被“稀释”；而对照组部分成员受实验组影响自主接受干预措施，可比喻为对照组被“污染”。无论是“稀释”还是“污染”均可使实验效应失去真实性。在实际工作中对“稀释”的补救办法，是将实验组成员按实际干预强度（剂量×时间）重新分组，然后与对照组去比较；若对照组成员被“污染”，可将被“污染”者去除后与实验组相比较。但是比较分析时，应注意两组的可比性是否被破坏。若失衡严重，对干预效果作因果解释时应慎重。

（收稿：1995-12-05）

青岛地区不同人群甲乙丙丁戊五型肝炎病毒感染情况的调查研究

张南霞[1] 赵春燕[1] 江崇才[2] 李运来[1] 苏乃伦[2] 宋永宁[1]

我们对青岛地区不同职业的部分人群进行了肝炎病毒的免疫血清学检测，现将主要结果报告如下。

一、对象与方法：共调查1007例（男575例，女432例），年龄18～60岁。其中肝病组：同期在中医院肝炎病房及肿瘤医院住院病人101例。根据1990年（沪）全国病毒性肝炎学术会议修定诊断分型标准定为慢性肝炎（CPH）22例，肝硬化（LC）66例和肝癌（HCC）13例；特殊职业组：储蓄员122例、长途司机172例、职业献血员160例；一般人群组：机关人员192例、食品加工人员132例，宾馆服务员173例，以上肝病患者及各职业人群均抽取空腹静脉血3～4ml，分离血清、置－20℃待检。进行抗-HAV-IgG、HBsAg、HBeAg、抗-HBe、抗-HBc、抗-HBc-IgM、抗-HCV、抗-HDV·HDVAg、抗-HEV 检测。试盒分别由上海科华实业公司和珠海亚利生物有限公司提供，均在有效期内。操作及结果判定由专人严格按试盒说明书进行。采用 ELISA 法。

二、结果：①抗-HAV、抗-HCV、抗-HDV、抗-HEV 检出率（%）、HBV 感染率（%），三组人群分别是：肝病组85.1、11.9、4.95、7.9、83.2；特殊职业组：长途司机89.0、0.79、0、2.36、46.5，储蓄员83.6、6.56、0、4.92、52.5，献血员80.6、6.25、0、3.13、8.1，一般人群组：食品加工79.5、0、0、0、25.8，机关人员80.7、0、0、0、22.9，宾馆服务员76.9、1.16、0、0.58、24.9。②HBV、HEV 以肝病组最高（83.2、7.9），储蓄员次之（52.5、49.2）。同其他人群比较，差异显著（$P<0.01$），HCV 以肝病组最高（11.9），其次为储蓄员（6.56）、献血员（6.25）、宾馆服务员（1.16）。机关人员及食品加工未查出。抗-HDV 仅在乙肝患者中查出5例，占4.95%。③肝病组 HBV 感染率最高尤以 CPH 活动期为甚（81.8%），HCV 感染以 HCC 最高，LC 次之，HDV、HEV 感染以 HCC 最高。重叠感染中以 HBV＋HCV 为最高（47.4，9/19），HBV＋HCV＋HEV 三重感染最低（5.2，1/19），抗-HDV 阳性均合并 HBV 感染。

三、讨论：①本次调查表明本地区各型肝炎病毒感染总阳性率以肝病组最高，其次分别是储蓄员、长途司机、宾馆服务员等，献血员最低。②各型肝炎病毒之间均可重叠感染，并能加重病情，加速肝炎向肝硬化或肝癌转化，肝癌及肝硬化中有12.7%未查出五型肝炎病毒标志物，不可忽视是否存在庚型肝炎病毒或其他致癌因素及国内尚不能检出的 GB 病毒感染。③储蓄员、长途司机感染率较高，与该职业日常工作环境及卫生条件有密切关系，应加强卫生管理。④献血员抗-HCV 阳性率较高且与献血次数呈正相关。抗-HEV 占3.13%，提示不可忽视 HEV 血液传播途径。⑤HDV 可重叠于 HBsAg 阴性的 HBV 感染。⑥抗-HAV·IgG 水平各人群普遍高，确定现症病人应结合临床查抗-HAV·IgM。

（收稿：1995-12-06 修回：1995-12-25）

1 山东省青岛市市北区卫生防疫站 266011
2 青岛市儿童医院