中华流行病学杂志1983年第4卷第1期

讲座

现场调查中常见的系统误差及其控制方法

北京医学院　王绍贤

现代医学的发展使医学家们日益认识到许多生理病理现象的出现，往往不是某一单一因子所能解释的，影响某一疾病发生发展和防治效果的因素，常常是众多因素错综复杂地交互起作用。为了描述许多现象的内在规律，医学家们除了在实验室开展细胞、分子的研究，在临床创立显微外科、器官移植等先进技术外，又都纷纷走向现场，开展了极为广泛的流行病学研究，如癌症、心血管病、计划生育等方面的现场流行病学研究，国内外均有不少报道。由于现场调查有其本身的特点及规律，如果掌握不好，往往导致较大的偏误，从而影响调查结果的准确性。为适应当前广泛开展现场调查研究的需要，很有必要对现场调查中误差的来源及其控制方法进行研究探讨，以助于提高现场调查的质量。

现场调查的种类及其特点

现场调查区别于实验室及临床研究的是因素复杂不易控制，许多调查都是在众多因素客观存在的情况下不加控制地进行的。但调查的目的不是消极地反映情况，任何一个调查研究都应有其具体的目的，如有的调查需要了解某个病的总的患病水平，有的调查要明确两个现象之间的关系等，因而有的调查要求面广些，有的调查则要求深入些。一般按照调查的深度广度不同，将现场调查分为几个类型。

一、普查（Census）：普查是对研究总体中的每个单元均进行调查，通常用来了解某一现象在总体中的分布，或总体中某现象的特征，这种调查的典型代表是人口普查。这类调查因耗费人力物力巨大，故一般不经常做，许多国家都采用每5～10年进行一次人口普查；另一方面，这类调查由于调查面广，工作量大，不可能进行较深入细致的调查。因此，对医学研究来说，普查不是一种理想的方法，但在某些情况下，如要对某病进行普治，就需要了解所有防治对象，以便采取防治措施，这就需要用普查的方法。

二、典型调查（Case Study）：典型调查也叫案例调查，这种调查和其他调查的一个主要区别是它的调查数是“1”，这个“1”可以是一个人、一个组织、一个社会等。旅行家对新大陆的描述、经济学家对某地经济发展与劳动资源的调查研究，医学家对某个疾病的流行病学特点的研究等均属于典型调查。这种调查最大的优点是可以进行较深入细致的研究，而且现场便于组织。但由于它调查的只是一个案例，既未考虑要研究的总体，也不是随机抽样，故无法计算抽样误差，不能从统计学上对总体进行推论，而研究者则往往希望自己的成果能推论更广一些，更概括一些，在某些情况下，当所调查的案例具有很大的典型性时，有时用典型调查推论总体也是可以的。例如，对北京市西城区的生育率调查能否推论北京市城区的生育率？这就取决于西城区育龄妇女的年龄、职业、文化程度、经济收入等影响生育率的特征的分布情况，对北京市城区的育龄妇女来说是否具有典型性，或者说是否具有代表性，如果具有代表性，那就可以用西城区的生育率来推论北京市城区的生育率，但这种推论是根据专业知识和经验判断来推论的，而不是统计学的推论。

三、抽样调查（Sampling Surrey）：所谓抽样调查是从研究总体中抽取一个样本进行调查，从样本的结果来认识推论总体。人们认识客观世界几乎都是通过抽样的方法来认识的，故抽样方法可以说是一种很古老的方法，但抽样方法发展成为一门学科，还是最近30年来的事。由于概率论的应用，使得用样本推论总体更具有理论依据和科学性。众所周知，用任何精确的样本去估计总体总是会有误差的，如果样本是随机抽样(也称为概率抽样)所得，就可以计算出样本中每个单元从总体中被抽出的概率，从而在用样本推论总体时，可以估算出误差大小及其发生的概率，这样就使得用较小样本推论较大总体有了依据。当然，这样做的先决条件是必须使用概率样本(Probability Sample)，即样本中每个单元从总体中抽出的概率是相等的，无论是简单随机抽样，系统抽样、分

层抽样抑或整群抽样，这一点都是必须遵守的原则，否则就不能计算抽样误差，也就不能从统计学上用样本推论总体。与概率样本相反的是非概率样本（Non-Probability Sample），如研究者在交通要道或公园里，观察过往行人有多少吸烟，多少不吸烟，尽管过往的人包括男女老少，三教九流，但无法计算他们每个人从总体中抽出的概率，换句话说，无法知道他们出现在街头或公园是否随机的，故算不出这种样本推论总体的误差及其概率。又如常见这样的样本——男女各抽100人，这往往也是非概率样本，因为研究总体中男、女各占的比例常常不是1：1的关系，故这种样本也不能计算用样本推论总体的误差及其概率。总之，只有概率样本才能从统计学上推论总体，非概率样本只能依据判断和经验对总体进行推论，这种推论必须非常小心。

现场调查中误差的种类及其作用

任何科学研究，不论其精密度有多高，总是会有误差存在的。现场调查由于因素众多，不好控制，产生误差更是不言而喻的。为了保证现场调查的质量，就必须尽量减少误差，因而有必要对现场调查中的误差进行较深入的研究，这方面国内外学者均做过不少工作。现场调查中的误差按其来源和性质不同可以分为两大类。

一、随机误差（Random error）或变量误差（Variable error）：这类误差主要来源于抽样误差，有时也包括一些随机测量误差，但主要是抽样误差，故通常把这类误差笼统地称为抽样误差（Sampling error）。这类误差存在于各种抽样方法的调查研究中，但普查中没有这类误差。在概率抽样中，由于抽样是随机的，故可以对抽样误差进行测量；而在非概率抽样中，由于抽样不是随机的，因而虽有抽样误差存在，但是不可测的。

二、系统误差（Systimatic error）或偏差（Bias）或非抽样误差（Non-Sampling error）：这类误差不但广泛存在于各类调查研究中（普查、典型调查及抽样调查），而且广泛存在于调查研究的各个环节，调查员诱导式的提问，被调查者的特定心理状态，以及调查表的设计质量等均可导致系统误差。系统误差虽然不像抽样误差一样可以测量，但它也是可以控制的，不过难度较大，原因是到目前为止，这类误差的控制方法还不是数学模型，而主要是靠高质量的调查表及调查员。

现场调查中，这两类误差控制不好，都会影响调查结果的准确性，但二者由于性质不同，其作用大小也不同，现在我们举例说明：A、B两个射手打靶，A射手技术好，但枪不好，故多次射击情况相差不大，但每次离靶心都很远。A所射的位置和靶心的距离为系统误差；B射手技术不好，但枪好，多次射击情况相差很大，但离靶心均较近，B所射的位置与靶心的距离为随机误差（图1）。由此可见：当有较大的系统误差存在时，即使随机误差很小，其结果偏误仍很大；相反的，当系统误差较小时，即使随机误差较大，其结果偏误并不大。我们还可以用图解(图2)来进一步说明这个问题。图2中横轴为样本均值，μ为总体均数所在的位置，纵轴代表样本数。从同一总体中多次抽样，若得到A、B、C、D四种样本均值的分布，结合总体均数μ的位置，我们可以对这四种分布得出结论：

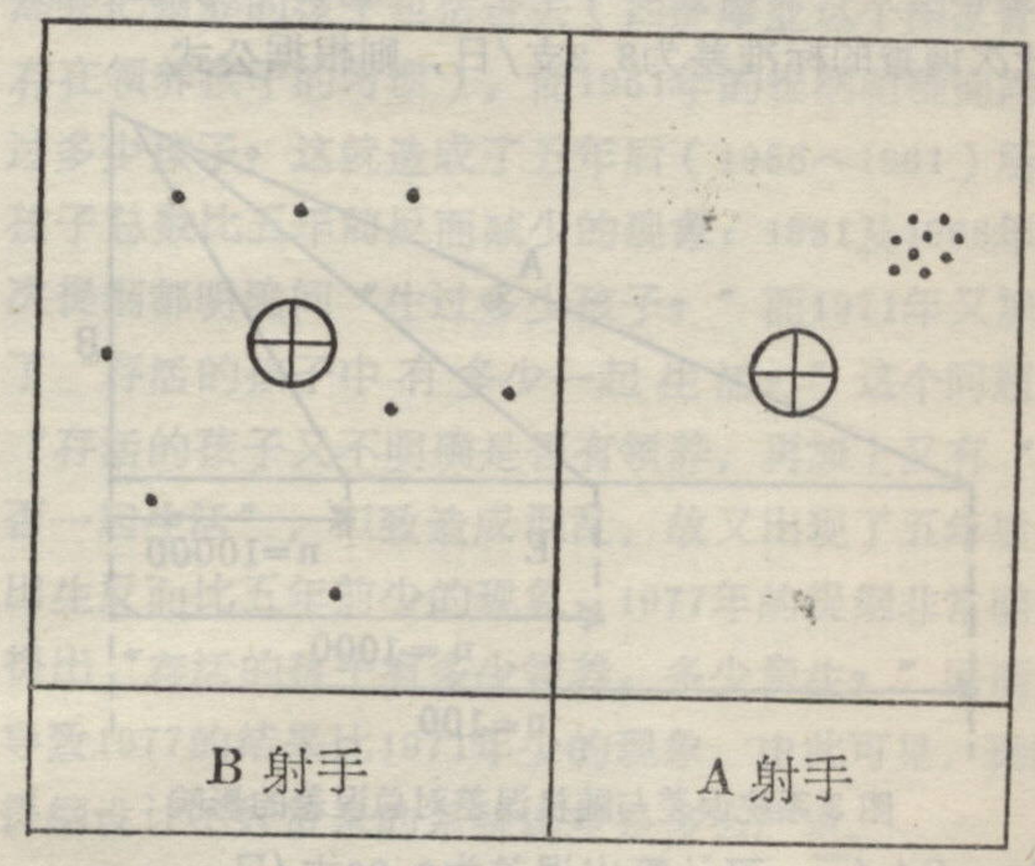

图1　两个射手的系统误差比较

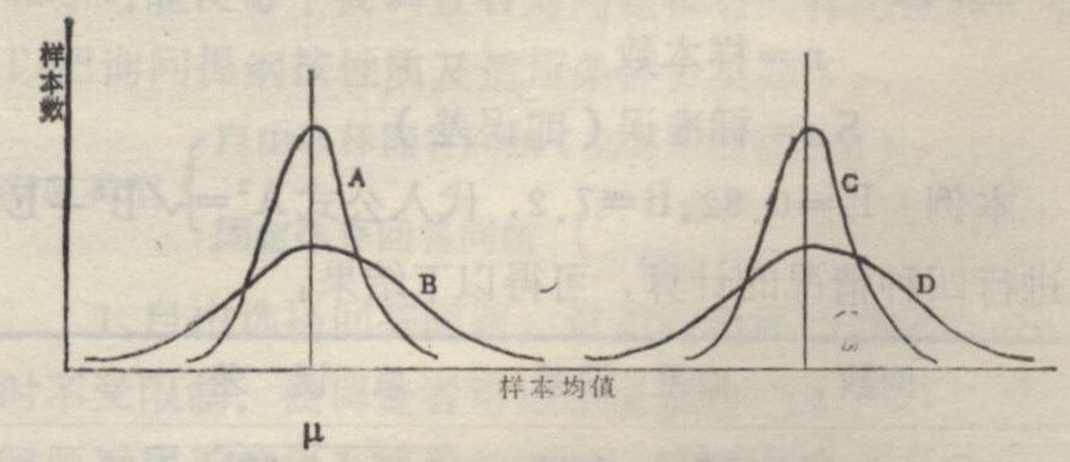

图2 系统误差与随机误差对总体均数的偏离

1.A分布样本均数较集中，而且靠近总体均数μ，说明随机误差及系统误差都较小。

2.B分布样本均值较分散，但还在μ周围，说明随机误差大，但系统误差小。

3.C分布样本均数虽然较集中，但远离总体均数

μ，说明随机误差虽然小，但系统误差大。

4.D分布样本均数较分散且远离总体均数μ，说明抽样误差大，系统误差也大。这个图解告诉我们，当系统误差很大时，随机误差再小也不能反映总体的真实情况。

随机误差的大小可以通过抽样技术（如分层可使层内变异减少）及增加样本数来控制，但各类系统误差却不能用增加样本数来控制。例如，已知某地吸烟的青工平均吸烟量为22.5支/日，现抽样调查100人，被调查者都愿意把自己的吸烟量说少一些，故平均值为15.3支/日，此时系统误差为22.5－15.3＝7.2支/日，若以直角三角形来表示两类误差的关系(图3)，其中：

A＝总误差

B＝系统误差（Bias）

E＝随机误差（errors）

则 $A^2=B^2+E^2$　即总误差＝$\sqrt{偏差^2+误差^2}$ 若设此次调查的标准差为8.2支/日，则根据公式

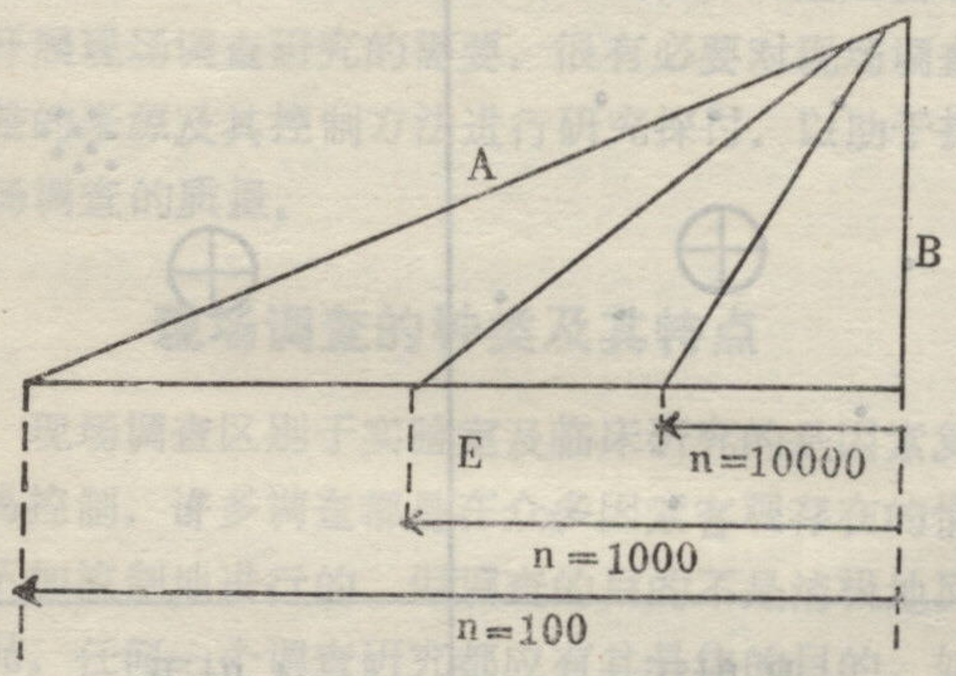

图3 系统误差与随机误差对总误差的影响

$S_{\bar{x}}=S/\sqrt{n}$　可计算出误差为0.82支/日

上式中　S＝标准差

n＝样本数

$S_{\bar{x}}$＝标准误（即误差）

本例　E＝0.82，B＝7.2，代入公式 $A^2=\sqrt{B^2+E^2}$ 进行四种情况的计算，可得以下结果：

	例数	偏差	误差	总误差
①	100	7.2	0.82	$\sqrt{7.2^2+0.82^2}=7.25$
②	1,000	7.2	0.26	$\sqrt{7.2^2+0.26^2}=7.20$
③	10,000	7.2	0.08	$\sqrt{7.2^2+0.08^2}=7.20$
④	10,000	14.4	0.08	$\sqrt{14.4^2+0.08^2}=14.4$

从①～③可看到：样本数加大至10倍，随机误差仅缩小3倍左右，加大至100倍，随机误差缩小10倍；系统误差不变，则样本数为100，1,000，及10,000时的总误差几乎相等；相反的，第④种情况，随机误差不变，系统误差只增加1倍，总误差也增加1倍。由于系统误差在数字上永远是相加的作用（随机误差则无一定方向，可以互相抵销），故加大样本数只能缩小随机误差，但有可能增加系统误差，而且因为加大样本数所增加的系统误差远比由于加大样本数所缩小的抽样误差要大得多。由此可见，实际工作中不应盲目追求样本数量，在考虑调查的误差时，必须同时兼顾随机误差及系统误差两个方面，而且必须把调查的精度和成本结合起来考虑。大家知道，在给定精度的情况下，可以由下式推算样本数。

$$n=\frac{t^2\ s^2}{d^2}$$

式中：n＝样本数

t＝设计要求的可信区间的t值（如95%的可信区间，当自由度＝∞大，t为1.96，此数值可按所要求的可信区间百分数，从正态分布面积表查到）

d＝容许误差（d＝t×标准误）

s＝标准差

若设容许误差为1，t为2，s为2时，

$$n=\frac{2^2\times 2^2}{1^2}=16$$

若设容许误差缩小1倍，即d＝0.5，t及s不变，则

$$n=\frac{2^2\times 2^2}{0.5^2}=64$$

可见容许误差缩小1倍，n要增加4倍，这样成本要大得多，更何况加大样本数还会加大系统误差，故盲目追求所谓“精度高，样本大”往往是得不偿失的。

在用样本推论总体时，也仍然是两种误差都应考虑。因为在现场调查中，总体所包含的范围是有不同含义的，我们可用以下图解来说明。

推论总体(Inference Population)→全中国7～14岁儿童

↑经验判断推论

目标总体（Target Population）→1981年7月全中国7～14岁儿童

↑经验判断推论

范围总体（Frame Population）→1981年7月除外台湾及西藏的中国7～14岁儿童

↑经验判断推论

调查总体（Survey Population）→在范围总体中又除外了住医院、外出、住拘留所等。

↑统计推论

样本（Sample）

假设我们要了解全国学龄儿童（7～14岁）的龋齿患病率，全中国7～14岁的少年儿童就应该是我们的推论总体，但在具体实施时，我们必须规定调查的时间，

如1981年7月调查，则1981年7月全国7～14岁的儿童就是我们的目标总体。根据调查目的具体规定一个目标总体还是比较容易的，但调查时往往又有某种情况使总体的范围有所变化，如我们可能没有调查西藏及台湾，这就是所谓的范围总体，在实际调查时，往往也不是范围总体，因为调查当时住医院的，到外地去的、在拘留所的等都未调查，故真正的调查总体比范围总体还小。严格来讲，从统计学上，样本只能推论调查总体，样本和调查总体之间的差是随机抽样误差，而再往上的总体和样本之间的差，已不是抽样误差，而是系统误差了（抽样的总体已不是所要推论的总体），故再往上推，就不属于统计学的范围，而是经验判断的推理。此时，如果系统误差较大，则不能随便往上推。

常见的系统误差及其控制方法

前面已提到，系统误差不仅存在于各种类型的调查中，而且存在于调查提纲、调查员及被调查者等环节中，现讨论如下。

一、询问提纲（调查表）所致的系统误差：现场调查中，主要的手段是对调查对象提出询问。询问提纲设计不当时，往往造成被调查者对同一问题发生不同的理解，使调查结果发生较大的偏误。例如，非洲国家西萨摩亚在每5年一次的人口普查中，都要询问妇女"一共生过多少孩子"，表1是1956～1976五次人口普查有关此问题的调查结果。同一年龄组的妇女所生的孩子数，随着时间的推移，应该愈来愈多，即第二次普查所生孩子数应该比第一次普查时多，即使在两次普查间，该组妇女一个孩子也没有生，两次普查所生孩子数也只是相等，绝不应该出现减少的现象，但表1中五次普查的结果，出现了许多后一次调查所生孩子数反而比前一次调查减少的数据（表中带*号者均是），问题出在什么地方呢？反回去找原始调查表时发现了问题，下面是五次普查有关此项目的询问提纲。

1956年　你共有多少孩子？存活多少？
1961年　你共生过多少孩子？存活多少？
1966年　你共生过多少孩子？存活多少？
1971年　你共生过多少孩子？存活的孩子中有多少和你一起生活？
1976年　你共生过多少孩子？存活的孩子中有多少是亲生的？有多少是领养的？

表1　西萨摩亚平均每个妇女所生孩子数

妇女年龄组（岁）	普查日期				
	1956	1961	1966	1971	1976
15-	.14	.12	.11	.09	.06
20-	1.31	1.33	1.43	1.14	.88
25-	2.94	3.13	3.38	3.15	2.71
30-	4.56	4.69	5.25	5.02	4.54
35-	6.03	5.84	6.48	6.17	5.90
40-	6.68	6.97	7.36	6.69	6.66
45-	7.19	6.98	7.77	6.76*	6.66*
50-	7.14	6.82*	7.27	6.62*	6.45*
55-	7.28	7.14	7.37	6.19*	6.51*
60-	7.19	6.96*	7.51	6.23*	5.96*
65-	6.89	7.10*	7.43	6.21*	5.87*
70-	6.31	6.58*	7.36	5.89*	5.90*
75+	6.44	6.36	7.27	5.64*	5.85*

编者注：此表内数字间原有斜线相连，即第一纵行的第一个数（.14）与第二纵行的第二个数（1.33）相连，再与第三纵行的第三个数（3.38）相连，其余数字连法依此类推，直部至将全表内数字连完，共42条短斜线。因排版困难，将其全取消，见谅。

从以上询问提纲可以看出：1956年的提纲不是明确问生过多少孩子？而问的是"共有多少孩子？"这就必然会把领养的孩子包括进去（西萨摩亚这个国家普遍存在领养孩子的习惯），而1961年的提纲明确询问生过多少孩子？这就造成了五年后（1956～1961）所生孩子总数比五年前反而减少的现象；1961及1966年两次提纲都明确问"生过多少孩子？"而1971年又加问了"存活的孩子中有多少一起生活？"这个问题中"存活的孩子又不明确是否有领养，再加上又有"是否一起生活"，以致造成混乱，故又出现了五年后的出生反而比五年前少的现象；1977年的提纲非常明确提出"存活的孩子有多少领养，多少亲生？"因而又导致1977的结果比1971年少的现象。由此可见，询问提纲设计不好带来的系统误差是多么严重。

为了便于设计询问提纲，使之尽可能的明确具体和尽可能使每个被调查者对问题都有同样的理解，可以把询问提纲按性质及适用条件分类如下：

问题类型
- 自由选择回答问题（回答不限制范围）
- 固定选择回答问题
  - 两种选择（"是""否"问题）
  - 多种选择

1.自由选择回答问题：这类问题被调查者在回答时不受限制，被调查者可以尽情谈自己的意见，这类问题适用于较复杂情况的调查，如调查避孕措施的效果常采用这种问题，它的优点是能了解到固定选择问题所不能了解到的情况。例如，要了解一个妇女所采用的避孕措施，如果用固定选择回答问题，就问："你是用口服药、避孕环还是安全期？"除此以外的一些避孕措施的了解就受到限制，但若用自由选择回答问题"你曾经用过什么避孕措施？"情况就大不一样，

如有的妇女在回答这个问题时谈到她曾用阿司匹林放在阴道内作为避孕药。另外自由选择回答问题易于拉家常、气氛轻松，易引起被调查者的兴趣，缺点是易出现“跑题”现象，调查较费时间，费用较大。

2.两种选择回答问题：两种选择回答问题是固定选择回答的一种，回答者只能回答“是”或“否”，“有”或“没有”，不能有第三种回答，如“你做过人工流产吗？”这种问题的优点是提问直接了当，最接近主题，回答简单，便于记录制表，缺点是只能满足两个相反的情况，若事实上只可能有两种回答时，所得结果就较准确，若事实与问题不符，如要了解妇女对人工流产的态度，问“你认为人工流产对妇女健康好不好？”只能回答“好”或“不好”，事实是一般情况下，人工流产对妇女健康不好，但某些情况，如因健康情况必须终止妊娠时，人工流产对妇女健康就是好的，因此这种问题就可以发生偏误。

3.多种选择回答问题：多种选择回答问题，避免了上述两种问题的缺点（不切题及只包括两种相反的回答），故被广泛采用，但它的重要缺点是很多情况下，被调查者必须认真思考后才能回答，容易造成被调查者精神上的紧张拘束，如医生问病人：“你们家每月要吃多少斤鸡蛋？吃到你口中每天能有多少？两个？一个？还是半个？”这种问题病人回答时就必须思考一番才行，此时，如果医生再一催问，则很容易随便回答一个数，如回答“就按1个算吧”。使用这种类型的问题时要注意几个问题：a、所给出的选择数量不能太少或太多，选择数量太少，则不成其为多种选择，选择数量太多则被调查者记不住，不好回答。如一连列举了十几种避孕措施，被调查者听了后面，忘了前面，回答就易发生偏误。b、在选择数量答案时，应注意排列顺序，如问“你认为中国目前的出生率是多少？”顺序列出10‰，15‰，20‰，25‰，30‰五个选择答案，调查结果回答20‰的人最多，并不是他们知道20‰是正确的，而是因为这个数字在中间，在这种情况下，排列顺序应该打乱。

实际工作中，这几种问题可以同时应用，或对一个问题采用不同类型的提问，以达到互补的作用。例如调查人们对人工流产的认识，第一步可以采用一个自由选择回答问题，以了解人们对人工流产的大致态度，如“你认为政府对人工流产应采取什么措施？”第二步可以通过一个两种选择回答问题来了解人们对人工流产的准确态度，如“如果有人建议禁止人工流产，你是赞成还是反对？”第三步可以通过多种选择回答问题，了解人们对人工流产赞成或反对的程度，如“你是非常拥护、赞成还是一般同意，还是反对禁止人工流产？”

二、调查员所致的系统误差：根据调查的目的及情况不同，对调查员的性别、民族等均应有不同的要求，否则往往会导致结果的偏误。例如，调查生育情况，最好用女调查员，在回族聚居地区最好不用汉族调查员，要想评价计划生育措施的效果，最好不用计划生育工作干部，否则，不易得到真实的结果。一个性别、民族…等都合适的调查员，也仍然会引起系统误差，这是因为调查者本身的思想、观点和态度常常能影响或控制被调查者。德国有一个心理学家曾做过这样一个实验，对同年级两个班的大学生分别介绍某教师的情况，对甲班说“下次给你们上课的是一位热情开朗、极为风趣的老师”；而对乙班说“下次给你们上课的是一位呆板冷淡，毫无趣味的老师”，然后同一个老师给两个班的学生上了同样内容的课，课后调查了两个班的学生对这位老师的反应，甲班反应积极，乙班反应消极。实际工作中调查员自觉或不自觉地对被调查者起作用的例子也是很常见的，如某地调查推行计划生育工作的效果，以群众对待“只生一个孩子的态度”为指标，调查时如果调查员像计划生育工作人员一样，边调查边宣传“只生育一个子女”的好处，则被调查者就容易回答“只生一个子女好，我早就领了独生子女证了”，至于她把独生子女奖金单独存入银行，准备有机会生第二胎时全部退赔的事，则只字不提，这样调查结果就反映不了真实的情况。

由此可见，调查员的质量在调查研究中占多么重要的地位，正因为如此，每个调查研究均应根据本身的特点、要求选择合适的调查员，然后进行培训，使每个调查人认识到自己在整个调查研究中所处的重要地位，不但要熟悉有关的业务，而且在调查时不应加入自己的观点，不应做任何假设，调查中要严格按照培训要求的规定和步骤去做。

三、被调查者所致的系统误差：现场调查中，被调查者所致的误差统称为无反应（Non-response），意指调查没有结果或结果明显错误，不能利用，其原因大致可以分为以下几类：

①调查不到：例如工作时间对双职工家庭的调查往往找不到人，解决的办法当然最好是摸清在家的规律，再去调查，当然再去调查就有个成本核算问题，美国有人做过计算，如去一次就调查到，每个调查费为100，去两次才调查到时，每个调查费就到112，如

去过五次才调查到，每个调查费就到250，由此可见事先对要调查的对象作些必要的了解，据此安排调查计划是非常重要的。

②特定环境的影响：现场调查中，在某一特定环境下，会使被调查者不据实回答问题，以致造成结果的偏误。例如1981年有人调查吸烟问题，正巧碰到全国烟酒提价，有的被调查者误认为调查的目的可能是给吸烟的低工资收入者补助，于是对吸烟量就尽量多报，造成结果偏误。又如人口普查有一项“亲生子女数”的调查，如果某户孩子是领养的，由于现实社会上总有人要指指点点说“××是领养的”，轻则弄得人家家庭不和，重的甚至家破人亡，因此一般领养孩子的都不愿说是领养的，而且不愿意让更多的人知道此事，在这种情况下，调查员如果比较策略，通过居民组织了解情况，或个别访问，同时注意职业道德，不使被调查者受害，则有可能了解到真实情况，如果调查员采取开会调查的方式，或当着孩子的面就问，那领养孩子的人一定会回答“是亲生的”。

③拒绝回答：现场调查中，有时会碰到被调查者拒绝回答的情况。例如有人研究第一胎人工流产对妇女健康的影响，其中未婚人工流产占一定的比例，这种情况调查时往往遭到拒绝。近年来，国外发展了一种用来调查敏感问题的技巧，从侧面来达到目的，具体做法如下：

调查员同时对被调查者提出两个问题，一个是要调查的敏感问题“你在最近12个月内是否做过人工流产？”另一个是毫无关系的非敏感问题“你母亲是不是南方人？”请被调查者从这两问题中随机选择一个回答（可用掷硬币等方法来选择），这个选择只有被调查者自己知道，而且不把这个选择告诉调查员，被调查者只要据实回答“是”或“否”，通过概率的计算，从侧面得到对敏感问题回答“是”的百分率。

随机选择回答问题的概率为：

选中敏感问题的概率为$P=0.5$

选中非敏感问题的概率为$1-P=0.5$

设 λ＝总的回答“是”的百分比（即调查中对两个问题回答“是”的百分比）

π_a＝抽到敏感问题的人回答“是”的百分比

π_b＝抽到非敏感问题的人回答“是”的百分比则总的回答“是”的百分比λ就应为：

$$\lambda=P\pi_a+(1-P)\pi_b$$

π_a：对敏感问题回答“是”的概率；π_b：对非敏感问题回答“是”的概率

对敏感问题回答“是”的百分比就应为：

$$\pi_a=\frac{\lambda-(1-P)\pi_b}{P}$$

上式中π_b很容易通过单独调查获得，λ为本次调查结果，P为0.5，故π_a即可算出。例调查100名25～30岁的青年妇女（包括已婚未婚），采用上述方法调查，总的回答“是”的百分比为60%；从这些青年妇女所在街道中，了解到该地区45岁以上妇女中，南方人所占比例为40%，即

$\lambda=0.60$　$P=0.5$　$\pi_b=0.40$

代入上式：

$$\pi_a=\frac{0.6-(0.5\times0.4)}{0.5}=\frac{0.4}{0.5}=0.8=80\%$$

即在此100名妇女中，抽到敏感问题回答“是”的百分比为80%。

π_b可以通过单独调查获得的理由是：对非敏感问题来说，一般都能按实际情况回答，因此极易获得。

在“无反应”的问题上，尽管可以采取上述一些措施，但仍然解决不了全部问题，如有人可能多次访问都找不到，有人始终拒绝回答问题，也不愿参与“随机选择回答敏感和非敏感问题”的调查。因此有人提出了另外一些纠正因“无反应”所致的误差的办法。

Cochran 等提出修正可信区间的方法。这个方法的基本思想是把研究总体分成“有反应”及“无反应”两个层，第一层中包括的单元都是“有反应”的，第二层中包括的单元都是“无反应”的

设 N＝总体数

N_1＝第一层的数

N_2＝第二层的数

则各层的权数（即各层的数在总体中所占份量）为：$W_1=N_1/N$

$W_2=N_2/N$

W_2实际就是总体中“无反应”者所占的比例。若从总体中随机抽样，所得资料实际是来自第一层的随机样本，而没有第二层的资料，因此样本均数距离总体均数之差为：

$$\bar{x_1}-\mu=\bar{x_1}-(W_1\bar{x_1}+W_2\bar{x_2})$$
$$=W_2(\bar{x_1}-\bar{x_2})$$

注：$\bar{x_1}-(W_1\bar{x_1}+W_2\bar{x_2})$

$$=\bar{x_1}(1-W_1)-W_2\bar{x_2}$$
$$=\bar{x_1}\left(\frac{N-N_1}{N}\right)-W_2\bar{x_2}$$
$$=\bar{x_1}(N_2/N)-W_2\bar{x_2}$$
$$=W_2(\bar{x_1}-\bar{x_2})$$

从上式中可以看到“无反应”者所占的比例是决定样

本与总体间偏差大小的重要因素。在计量资料中，有了W_1及W_2尚不能有助于调整可信区间。而在计数资料中，由于第二层的均值P_2只会在0～1之间（即由无人回答至全部都回答），如果知道W_2，则可以调整可信区间。具体方法如下：

若从N_1中抽取样本n_1，得到P_1，当n_1足够大时，P_1的95%的可信区间为$P_1 \pm 2\sqrt{\frac{P_1(1-P_1)}{n_1}}$而总的样本P的可信区间应为：

W_1（P_1的可信区间）+W_2（P_2的可信区间）

而P_2的可信区间最安全的就是0—1，即以0为P_2的可信区间的下限，以1为P_2的可信区间的上限，则P的可信区间即为：

$\hat{P}1$（下限）$=W_1(P_1-2\sqrt{P_1(1-P_1)/n_1}+W_2(0)$

$\hat{P}u$（上限）$=W_1(P_1+2\sqrt{P_1(1-P_1)/n_1}+W_2(1)$

例　对1,000名青工进行吸烟调查，有反应的为850人，无反应的为150人，在有反应的850人中，吸烟者为320人，则

$P_1=320/850\times100\%=37.6\%$

$W_1=850/1000=0.85$

$W_2=150/1000=0.15$

若此1,000名青工为所有青工中的一个随机样本，则青工中吸烟者的比例P的95%的可信区间为：

$$\hat{P}1=0.85(0.38-2\sqrt{0.38(1-0.38)/850})+0.15(0)=0.29$$

$$\hat{P}u=0.85(0.38+2\sqrt{0.38(1-0.38)/850}+0.15(1)=0.50$$

即P的95%可信区间为29～50%

表2即为用上述方法计算所得的95%可信区间。

“无反应”的问题所致的误差是系统误差，即使抽样很好，也无助于解决“无反应”的问题，相反地，如果样本中“无反应”的多了，则会破坏原来抽样的随机性，所以人们纷纷探索解决“无反应”问题的办法。Kish提出，对调查不到的单元，可以从样本中随机抽取相应数量的单元，复制其结果，补充到样本中去。此法简单易行，但应注意“无反应”的单元是否有偏性，只有“无反应”的单元偏性不大时，用此法才较可靠。

表2　　95%可信区间　　n＝1,000

无反应者所占% W_2X100	样本百分数 P_1X100 5	10	20	50
0	(3.6,6.4)	(8.1,11.9)	(17.5,22.5)	(46.7,53.2)
5	(3.4,11.6)	(7.6,16.3)	(16.5,26.5)	(44.4,55.6)
10	(3.2,15.8)	(7.2,20.8)	(15.6,30.4)	(42.0,58.0)
15	(3.0,20.5)	(6.8,25.2)	(14.7,34.3)	(39.6,60.4)
20	(2.8,25.2)	(6.3,29.7)	(13.7,38.3)	(37.2,62.8)

四、抽样编码所致的系统误差：在应用随机抽样时，要事先将总体中的单元编码登记(即编制名册)，然后从中随机抽取所需的样本数。编码时，要求单元与编码有一一对应的关系，不得有重复、遗漏等现象发生，若设L为编码，S为单元，则二者应有如下的对应关系：

```
S   S   S   S
|   |   |   |
L   L   L   L
```

实际工作中，由于总体包含的单元很多，难免不发生这样那样的问题，从而破坏了L与S的一一对应关系，这就是抽样中所谓的“范围”(Frame)问题，这类问题处理不当，就会破坏原来抽样设计的随机性，形成了不等概率抽样。常见的Frame问题如下：

①缺失单元

```
S   S   O   S   O
|   |   |   |   |
L   L   L   L   L
```

如死亡、迁移等。

②编码重复

```
S       S       S
|     /   \     |
L   L   L   L   L
```

如一个人有数个住处，按住处编码即有此现象。

③一个编码代表数个单元

```
S     S   S   S     S
|       \   /       |
L         L         L
```

如以宿舍楼为单位编码，而抽样时以户为单元抽，就会出现几个户共有1的编码

④遗漏编码

```
S   S   S   S
|   |   |   |
L   O   O   L
```

如按电话薄编码，有的单元电话薄上没有。

以上问题主要应在编码前做好准备工作，如先到有关的居委会核实了解情况，然后再编码，这样多数的Frame问题均可避免，实在避免不了，则按情况不同采取补救办法。如缺乏单元可按“无反应”处理；编码重复可将重复者作为零值处理；一个编码代表数个单元时，可先用整群抽样后，在抽到的整群中再抽单元等办法来处理。