1. When to do and when not do it.

某公司已经有了一个免费的APP，现在公司想上线新功能，但这需要用户升级软件、进行注册才能体验新功能。这个不能用AB测试，因为这种需要用户主动做一些事情才能体验到新功能，所以没办法随机分配两组等量的用户然后进行测试（没办法设置很好的对照组）。你可以通过AB测试得到部分结果，如用户是否愿意进行注册（以体验新功能），但你无法通过AB测试完成所有的检验（如新功能是否对公司有好的促进）。

大概就是说如果可以很清晰的分出实验组和对照组，硬件上可以执行AB实验，就可以了。

不好用AB Testing的时候：1。新的体验，用户对更换的新的体验表示很惊奇，或者有novelty detection。2.时间较长，不好测定的指标。比如租房屋的APP，推荐朋友租房子，人们并不是非常经常浏览租房子的系统或经常进行换租房活动，上线一个新的推荐系统，想要看到效果需要很长时间，这是AB testing不支持的。后面第三章会提到替代的解决办法。3.AB Testing不能告诉你你是否miss了一些事情。比如一个网页上面放了相机的一些评论，AB测试可以告诉你对于任两个评论放上去会更好，但没法告诉你是否漏掉了一些评论，哪些评论是应该补充上去的。4.公司更新品牌名称或者换了一个logo。这个不好用AB测试，因为用户情感波动太强了，换logo是一个很大的事情，需要时间去适应，不能短期实验得到结果，而且也不好定义具体的metric。

一些好的AB测试的例子：

Google tests 41 different shades of blue.

Metric choice

场景是一个在线教育平台，主页面有个 start now 按钮，想更改它的颜色。选择metric。最先想到的是学生完成的总课程数量，因为这是我们想直接提升的因素（增加课程销量，提高销售额）。但这有个问题就是学生可能需要比较久的时间去完成上课，而这不适合AB测试-----一个短期完成的测试。然后想到的可能是用户点击start now按钮的次数。因为从漏斗分析来讲，这样做可以增大漏斗后面层次的用户数。但这也会有问题，因为次数是一个相对不稳定的东西，会受到很多因素影响，如外部降价因素，环境因素等（这句话是我自己理解的）。还有一个大问题是，我们分配给对照组和实验组的总人数是一定的，但这些人里面有部分会浏览主页面，再其中部分去点击start now按钮等等，所以实验组和对照组，分配的总人数是一样的，但是点击主页面的人数不一定一样，点击start now按钮的人也不一定一样。如对照组100个人点击了主页面，20个人点击了按钮，对照组10个人点击了主页面，8个人点击了按钮。实验组点击按钮的用户数是20，大于对照组8，但可能对照组的效果更好，因为比率更高。那么用什么比率指标去衡量呢？CTR click through rate。# of clicks/# of page views.这个指标的问题是它可能大于1.比如有两个visitor，一个没点按钮，一个点了五次（可能网页load的慢他不耐烦），这样的CTR就是5/2=2.5.所以我们用CTR probability，which is: # unique visitors who has clicked the button at least once / # unique visitors who has clicked the main page.