

# 真实世界研究样本量估计的统计学考虑

翟静波  郑文科  王  辉  金鑫瑶  张俊华  
(天津中医药大学循证医学中心,天津,301617)

**摘要**  真实世界研究包括多种设计类型。当受试对象来源于总体中的部分样本时,需要考虑样本量的估计问题。本文探讨了真实世界研究样本量估计的基本原则、分析了其影响因素,提供了几种常用的真实世界研究样本量估计方法,并对真实世界研究样本量估计中参与人员的角色进行分析,为研究者开展真实世界研究的样本量估计提供一定的方法学参考。  
**关键词**  真实世界研究;样本量;统计学;基本原则;影响因素;临床试验;队列研究;病例对照研究

## Statistical Considerations in Estimating Sample Size of Real World Study

Zhai Jingbo, Zheng Wenke, Wang Hui, Jin Xinyao, Zhang Junhua

(Evidence-based Medicine Center, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China)

**Abstract**  Real world study includes multiple design types. When subjects are derived from a sample in the population, estimating sample size needs to be considered. This study discussed the basic principles of sample size estimation in real world study, analyzed its influencing factors, provided several commonly used sample size estimation methods, and analyzed the roles of participants for estimating sample size in real world study. It provided a methodological reference for estimating sample size in real-world study.

**Key Words**  Real world study; Sample size; Statistics; Basic principle; Influence factor; Clinical trial; Cohort study; Case control study

中图分类号:R2-03    文献标识码:A    doi:10.3969/j.issn.1673-7202.2019.12.006

真实世界研究(Real-world Study, RWS)是指基于真实世界数据(Real-world Data, RWD)开展的研究<sup>[1-3]</sup>。尽管数据来源于真实世界,但是,研究者不太可能将所有符合条件的真实世界数据全部用于拟开展的研究。所以,在开展 RWS 时,研究者仍需考虑样本量的估计问题<sup>[4]</sup>。真实世界数据和大样本量之间并不能划等号。基于有限总体开展的 RWS,有可能是小样本的研究。由于数据来源于真实世界的特点、研究设计和研究目的多样性、经费来源的广泛性等,不同真实世界研究的样本量往往差异较大(几十例到上百例均有可能)<sup>[5]</sup>。如何合理估计样本量对于 RWS 的顺利实施至关重要。本文对 RWS 样本量估计过程中可能涉及的统计学问题进行了分析和讨论。

### 1 RWS 样本量估计的基本原则

总体(Population)是指根据研究目的确定的、具有同质性的所有观察单位构成的全体或集合,可以是所有人群,也可以是所有人群的某个测量指标的集合。总体包括有限总体和无限总体。传统的随机对照试验(Randomized Controlled Trial, RCT)的纳入和排除标准较为严格<sup>[6]</sup>。一般情况下, RCT 中纳入

的受试者是某一总体人群中的部分样本。RWS 的受试者来源于真实世界环境。这使得 RWS 有可能纳入某一总体人群进行研究。所以,在 RWS 设计阶段,研究者需要考虑受试对象来自于某一总体人群还是总体中的代表性样本。表 1 列出了常见 RWS 设计中受试对象的一般来源。

表 1  常见 RWS 设计中受试对象的一般来源

| 常见 RWS 设计类型      | 受试对象的一般来源 |    |
|------------------|-----------|----|
|                  | 总体        | 样本 |
| 实用性临床试验          |           | √  |
| 现况研究(普查)         | √         |    |
| 现况研究(抽样调查)       |           | √  |
| 病例对照研究           |           | √  |
| 基于有限总体的队列研究      | √         |    |
| 基于无限总体或总体未知的队列研究 |           | √  |
| 注册登记式研究          | √         | √  |
| 基于电子健康档案的研究      | √         | √  |

如果拟开展的 RWS 所纳入的受试对象是某个有限总体,那么,该 RWS 可以不用计算样本量,而是将这个总体人群均纳入研究。常见于普查、基于某个国家或地区全人口的队列研究等。

如果拟开展的 RWS 所纳入的受试对象所在的

有限总体过大或来自于无限总体,研究者无法获取总体。此时,研究者可以从总体中抽取部分有代表性的样本,通过样本来估计总体特征。此时,抽取样本量的多少需要通过样本量估计公式计算获得。常见于实用性临床试验、抽样调查、病例对照研究等。

当 RWS 的数据来源于一个或多个数据库时,研究者需要分析判断这些数据库纳入的人群是否代表了计划研究的总体人群。当 RWS 使用某一国家、地区人口或疾病数据库时,这类数据库在一定程度上代表了该国家或地区人口或疾病发展的趋势,可以考虑不计算样本量。但是,如果只是抽取其中有代表性的样本开展研究时,仍需考虑估计样本量。或者,基于可获得的人力、财力、物力,估计能够收集的最大样本量,通过计算把握度,评估有多大把握度能够获得事先假设的研究结果。常见于注册登记研究、基于电子健康档案的研究等<sup>[7-10]</sup>。

## 2 RWS 样本量估计的影响因素

**2.1 研究设计类型** RWS 的研究设计类型包括观察性研究和试验性研究<sup>[11]</sup>。观察性研究包括横断面研究、病例对照研究、队列研究等<sup>[11-12]</sup>。试验性研究包括单臂临床试验、实用性临床试验等<sup>[13]</sup>。不同研究设计所采用结局指标可能存在差异,所以样本量估计方法可能不同<sup>[14]</sup>。

**2.2 资料类型** 不同的结局指标分属于不同的资料类型。RWS 的资料类型包括定量资料和定性资料<sup>[15]</sup>。资料类型的不同,选用的样本量估计公式不同<sup>[16]</sup>。在一定条件下,定量资料与定性资料可以相互转换。所以,明确结局指标对应的资料类型是进行样本量估计的重要前提。

## 2.3 样本量估计参数<sup>[17]</sup>

**2.3.1 检验水准** 检验水准  $\alpha$  是指研究者在推断不同组间是否有统计学差异时所采用的概率水平。取值越小,所需样本量越多。一般情况下,检验水准  $\alpha$  的取值为 0.05<sup>[18]</sup>。

**2.3.2 检验效能** 检验效能也称为把握度,通过 1 减去 II 型误差 ( $\beta$ ) 获得。把握度越高,所需样本量越大。一般情况下,把握度的取值为 80% ~ 90%<sup>[19]</sup>。

**2.3.3 变异程度** 变异程度是指总体内部不同个体之间或总体之间的差异大小,常采用标准差估计。一般情况下,变异程度越大,所需样本量越大。

**2.3.4 单、双侧检验** 在其他参数不变的情况下,与双侧检验比较,单侧检验所需样本量更少。单侧和双侧检验的选择主要是根据研究目的和临床专业

角度考虑确定。

**2.3.5 脱落和失访** 在传统的随机对照试验样本量计算过程中,脱落和失访率通常设置为 10% ~ 20%。在真实世界环境下,受试者的脱落和失访问题可能更为突出。在确保研究质量的前提下,研究者可以在估计 RWS 样本量时适当增大脱落和失访率,从而使脱落和失访率符合真实世界的实际情况。

## 3 RWS 样本量估计的常用方法举例

PASS 软件是样本量估计的专业软件。本文以 PASS 软件为平台,简要介绍 RWS 中常用的样本量估计方法。

### 3.1 2 个独立样本定量资料比较的样本量估计

假设:研究者拟开展一项实用性随机对照试验。主要结局指标为收缩压下降值。观察组收缩压下降值为 10 mm Hg,对照组收缩压下降值为 7 mm Hg。2 组收缩压下降值的标准差均为 10 mm Hg。检验水准为 0.05。把握度为 90%。2 组样本量相等。如图 1 所示,将参数输入对应位置后,点击“RUN”,可得每组样本量为 235 例。考虑 10% 的脱落率之后,每组需要 259 例。

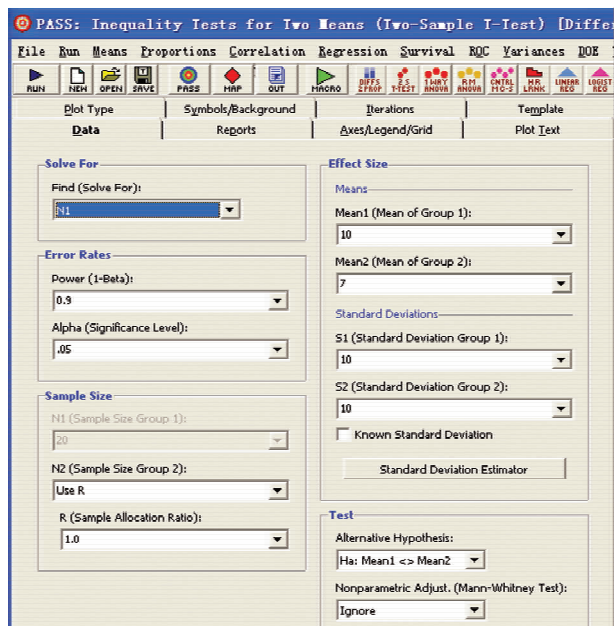


图 1 2 个独立样本定量资料比较的样本量估计举例

### 3.2 2 个独立样本定性资料比较的样本量估计

假设:研究者拟开展一项基于真实世界数据的队列研究。主要结局指标为某疾病发生率。暴露组疾病发生率为 9%。非暴露组疾病发生率为 7%。检验水准为 0.05。把握度为 90%。2 组样本量相等。如图 2 所示,将参数输入对应位置后,点击“RUN”,可得每组样本量为 3 865 例。考虑 10% 的脱落率之后,每组需要 4 252 例。

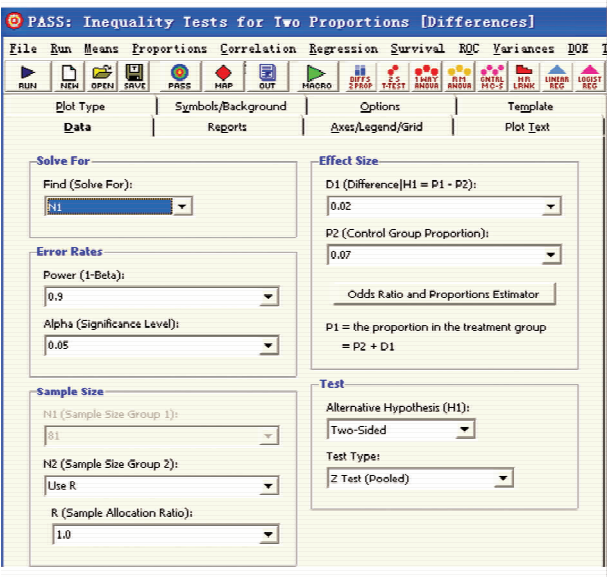


图 2 2 个独立样本定性资料比较的样本量估计举例

3.3 以中位生存时间为结局指标的样本量估计

假设:研究者拟开展一项基于真实世界数据的队列研究。主要结局指标为某疾病的中位生存时间。暴露组人群的中位生存时间为 20 个月。非暴露组人群的中位生存时间为 24 个月。检验水准为 0.05。把握度为 90%。2 组样本量相等。失访率为 10%。研究期限为 36 个月。受试对象入组期限为 12 个月。如图 3 所示,将参数输入对应位置后,点击“RUN”,可得每组样本量为 2 786 例。

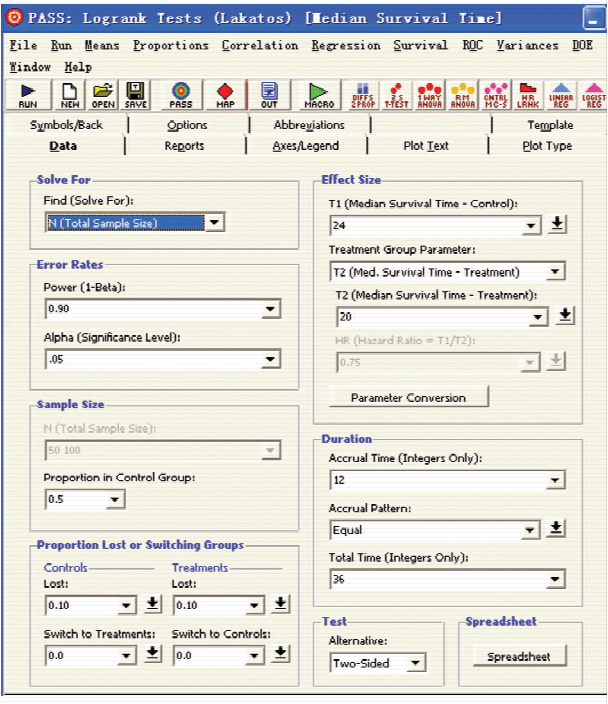


图 3 以中位生存时间为结局指标的样本量估计举例

3.4 以罕见事件发生率为结局指标的样本量估计  
在以观察某一人群中罕见事件发生率为结局指

标的研究(例如:药物安全性集中监测研究)中,一般不设置对照,只是纵向观察该人群某罕见事件的发生率。此时,我们可以假设:该罕见事件的发生服从二项分布或 Poisson 分布<sup>[20]</sup>。采用 SAS 软件编程获得相应的样本量估计结果。从表 2 和表 3 中可以看出,罕见事件服从二项分布和 Poisson 分布的样本量估计结果相近。从表 2 中可知,假设:罕见事件的实际发生率为 0.01%,如果研究者想以 95% 的把握发现至少 1 例罕见事件,则调查的人数为 29 956 例(约为 3 万例)。我们常见到一些药物安全性集中监测研究的样本量设定为 3 万例,其目的就是为了有 95% 的把握能够至少发现 1 例发生率为 0.01% 的不良反应<sup>[21-22]</sup>。

表 2 基于二项分布罕见事件发生率的样本量估计

| 罕见事件的<br>发生率 | 罕见事件的发现率 P(至少发现 1 例) |        |        |        |        |
|--------------|----------------------|--------|--------|--------|--------|
|              | 80%                  | 85%    | 90%    | 95%    | 99%    |
| 1%           | 161                  | 189    | 230    | 300    | 460    |
| 0.10%        | 1 609                | 1 897  | 2 300  | 2 995  | 4 600  |
| 0.01%        | 16 094               | 18 971 | 23 025 | 29 956 | 46 050 |

表 3 基于 Poisson 分布罕见事件发生率的样本量估计

| 罕见事件的<br>发生率 | 罕见事件的发现率 P(至少发现 1 例) |        |        |        |        |
|--------------|----------------------|--------|--------|--------|--------|
|              | 80%                  | 85%    | 90%    | 95%    | 99%    |
| 1%           | 161                  | 190    | 230    | 300    | 461    |
| 0.10%        | 1 609                | 1 897  | 2 303  | 2 996  | 4 605  |
| 0.01%        | 16 094               | 18 971 | 23 026 | 29 957 | 46 052 |

3.5 RWS 的把握度估计 当无法获取计划的样本量时,研究者可以预估一下可获得的、最多的样本量。利用这个样本量进行把握度估计<sup>[23-24]</sup>。

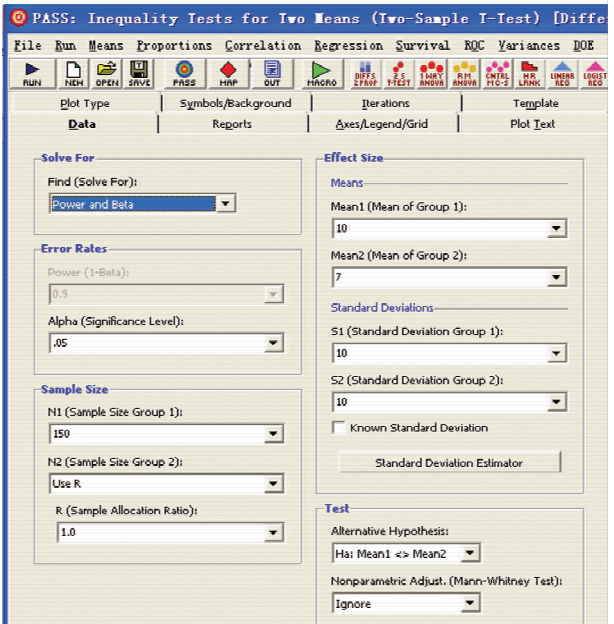


图 4 2 个独立样本定量资料比较的把握度估计举例

假设:研究者拟开展一项实用性随机对照试验。主要结局指标为收缩压下降值。观察组收缩压下降值为 10 mm Hg,对照组收缩压下降值为 7 mm Hg。2 组收缩压下降值的标准差均为 10 mm Hg。检验水准为 0.05。2 组样本量相等。研究者能够获得的多样本量为每组 150 例。如图 4 所示,将参数输入对应位置后,点击“RUN”,可得把握度为 74%。也就是说,研究者有 74% 的把握度识别出组间收缩压下降均值之差至少为 3 mm Hg 的差异。研究者可以根据这个把握度评估样本量是否足够支撑研究目的的实现。

#### 4 RWS 样本量估计中参与人员的角色

RWS 的样本量估计过程中需要多方人员参与,例如:临床专家、统计学专家、流行病学专家、申办方等。临床专家的主要任务是从临床实践角度出发,考虑最合适的结局指标和临床参数。统计学专家的主要任务是从统计学角度出发,考虑样本量估算过程中统计参数设置的合理性。流行病学专家的主要任务是从流行病学角度出发,考虑研究设计和相关参数的合理性。申办方的主要任务是基于所掌握的研究资源考虑样本量估算的可操作性。

#### 5 小结

RWS 设计类型多样,既有传统的流行病学研究设计,也有基于现代信息获取技术的健康相关档案数据研究。传统的流行病学研究设计,例如:RCT、病例对照研究、队列研究等,通常需要考虑样本量的估计问题。但是,基于注册登记、电子健康档案等开展的 RWS,可能会产生一些新的研究设计类型和统计分析方法。相关的样本量估计方法可能也将随之更新和完善。我们也期望这些新的设计和方法能够提高 RWS 的研究效率。

#### 参考文献

- [1] 孙鑫,谭婧,唐立,等. 重新认识真实世界研究[J]. 中国循证医学杂志,2017,17(2):126-130.
- [2] 姚晓莹,张靖雪,詹思延. 真实世界证据与随机对照试验:RCT DUPLICATE 项目概述[J]. 药物流行病学杂志,2019,28(8):495-517.
- [3] 张雨雨,陈静,赵晨,等. 真实世界理念下观察性与实验性临床研究协同应用[J]. 中国循证医学杂志,2018,18(4):284-288.
- [4] 曹越,尹庆锋,曾宪涛. 真实世界研究概述[J]. 武警医学,2017,28(4):400-403.
- [5] 黄卓山,罗艳婷,刘金来. 真实世界研究的方法与实践[J]. 循证

医学,2014,14(6):364-368.

- [6] 崔伟峰,王素花,王会丽,等. 真实世界中医临床研究方法与实践[J]. 中国全科医学,2016,19(13):1586-1597.
- [7] 孙鑫,谭婧,王雯,等. 建立真实世界数据与研究技术规范,促进中国真实世界证据的生产与使用[J]. 中国循证医学杂志,2019,19(7):755-762.
- [8] 王雯,谭婧,于川,等. 基于中国医院电子病历数据的药品安全性评价模式探索[J]. 中国药物警戒,2019,16(3):134-138.
- [9] 李晓春,戴国华,刘小翠. 基于 HIS 真实世界研究方法及其在中医临床疗效评价中的应用[J]. 山东中医药大学学报,2016,40(5):415-418.
- [10] 谭婧,程亮亮,王雯,等. 患者登记研究的策划与患者登记数据库构建:基于观察性设计的真实世界研究[J]. 中国循证医学杂志,2017,17(12):1365-1372.
- [11] Jing-Bo Zhai, Jiang Li, Jing Chen. The applications of Bayesian models in real-world studies of traditional Chinese medicine: a primer[J]. Traditional Medicine Research, 2017, 2(2):88-93.
- [12] 彭晓霞,舒啸尘,谭婧,等. 基于真实世界数据评价治疗结局的观察性研究设计技术规范[J]. 中国循证医学杂志,2019,19(7):779-786.
- [13] 唐立,康德英,喻佳洁,等. 实效性随机对照试验:真实世界研究的重要设计[J]. 中国循证医学杂志,2017,17(9):999-1004.
- [14] CCTS 工作组,陈平雁. 临床试验中样本量确定的统计学考虑[J]. 中国卫生统计,2015,32(4):727-733.
- [15] 胡良平. 医学统计学基础——资料类型的识别[J]. 中国骨肿瘤骨病,2002,1(4):227-229.
- [16] Charan J, Biswas T. How to calculate sample size for different study designs in medical research[J]. Indian J Psychol Med, 2013, 35(2):121-126.
- [17] 田峰,付莹坤,谢雁鸣. 中药上市后临床再评价试验样本含量估算方法探讨[J]. 中国中药杂志,2011,36(8):1097-1102.
- [18] 姚晨,阎小妍. 临床研究设计中的统计学考虑要点[J]. 北京大学学报:医学版,2010,42(6):632-636.
- [19] 杨进波. 探讨药物临床试验中样本含量及相关问题[J]. 中国临床药理学杂志,2007,23(1):68-69.
- [20] 杨进波. 从安全性评价角度探讨药物临床试验的样本含量[J]. 中国临床药理学杂志,2007,23(6):474-475.
- [21] 贺佳. 临床试验中药物安全性的统计学考虑[J]. 世界科学技术—中医药现代化,2017,19(7):1089-1096.
- [22] 郑青山,孙瑞元,陈志扬. 新药临床试验最低例数规定的安全性评价[J]. 中国临床药理学与治疗学,2003,8(3):354-355.
- [23] 高培,王杨,罗剑锋,等. 基于真实世界数据评价治疗结局研究的统计分析技术规范[J]. 中国循证医学杂志,2019,19(7):787-793.
- [24] 成小如,严若华,伯坚,等. 中国前瞻性城乡流行病学研究的设计与方案[J]. 中国循环杂志,2016,31(11):1088-1092.

(2019-11-10 收稿 责任编辑:徐颖)