

论文科研用户 文献信息分析和服

LUNWEN KEYAN YONGHU WENXIAN XINXI FENXI HE FUWU YANJIU

林佳瑜 著



华南理工大学出版社
SOUTH CHINA UNIVERSITY OF TECHNOLOGY PRESS

论文科研用户 文献信息分析和服 务研究

林佳瑜 著



华南理工大学出版社
SOUTH CHINA UNIVERSITY OF TECHNOLOGY PRESS

· 广州 ·

图书在版编目 (CIP) 数据

论文科研用户文献信息分析和 务研究/林佳瑜著. —广州: 华南理工大学出版社, 2018. 11

ISBN 978 -7 -5623 -5832 -9

I. ①论… II. ①林… III. ①科技文献 - 文献检索与利用 - 研究
IV. ①G254. 97

中国版本图书馆 CIP 数据核字 (2018) 第 243841 号

Lunwen Keyan Yonghu Wenxian Xinxi Fenxi He Fuwu Yanjiu

论文科研用户文献信息分析和 务研究

林佳瑜 著

出 版 人: 卢家明

出版发行: 华南理工大学出版社

(广州五山华南理工大学 17 号楼, 邮编 510640)

<http://www.scutpress.com.cn> E-mail: scutc13@scut.edu.cn

营销部电话: 020 - 87113487 87111048 (传真)

责任编辑: 卜穗珍

印 刷 者: 虎彩印艺股份有限公司

开 本: 787mm × 1092mm 1/16 印张: 11.25 字数: 196 千

版 次: 2018 年 11 月第 1 版 2018 年 11 月第 1 次印刷

定 价: 39.80 元

版权所有 盗版必究 印装差错 负责调换

前言

不同用户群体的科技论文下载和引用有许多不同的特征，文献与科研用户的关系值得深入挖掘。这其中涉及期刊、论文、图书等资料的使用情况，更与用户自身的学术修养、信息素养、检索技巧、检索偏好等密切相关，此外，与文献数据库自身的设计和功能的有着不可分割的关系。本书将通过5个章节展开详细探讨。

文献下载和引用次数与阅读使用次数的研究需要有针对性地、持续性地对科研用户下载和使用的资料进行详细数据的记录和跟踪调查，这是一个严谨、客观的过程。同时，进行这样一个专题研究是必要的，因为科研用户的下载和引用行为除了与自身的检索技巧有关外，更多的是与数据库后台的检索分析密切相关。对其进行定量分析既可对用户的检索行为做出评价，还可对数据库预设的检索条件进行评价，对日后研究数据库的智能检索及深入开发有重要意义。

文献的下载次数与阅读次数呈高度正相关是相对某些群体或个人而言的。文献的下载次数与阅读次数并不体现高度正相关在科研用户的检索行为上也能得到解释和论证。当然，科研用户的检索行为与论文下载次数和阅读使用次数的关系研究还有待更进一步深入调查分析，本书只是从用户的表象定量分析用户群体的差异以解释论文下载次数与阅读次数没有体现高度正相关，深层次的关联还需继续深入跟踪调查。此外，用户的检索行为与数据库的智能化关系密切，尽管数据库对每篇论文设置了同样的搜索概率，但是对于文章本身而言，除了作者和单位是编辑中确定的因素，其余的题名、

摘要甚至全文都是编辑过程中可变的因素。数据库如何利用这些可变的因素为用户提供更切合需求的论文，这也是需要进一步研究开发的。

下载和引用的发生与学者的专业知识水平层次有关。具备高专业知识修养（high-knowledge）的学者在下载和引用时会有选择性地偏向权威刊物和权威作者的文章，目光较多聚焦在近些年发表的文献，此外，他们的论文参考文献条数多，形式多样，参考的深度和广度更大；而专业知识水平相对较低（low-knowledge）的学者对下载和引用的文献选择门槛较低，要求较少，主要以便于获取的相关内容为主，对权威刊物和权威作者以及近期的研究文献关注度都相对较低，因而他们的参考文献集中度相对较低，论文的参考文献条数较少，引用范围较为狭小、单一。

高被引和高下载的论文大多数为专业知识较丰富的学者所创作，论文之间有较多重复的流行性和宽泛性的关键词出现，研究内容紧跟时代步伐，关注热门技术或理论，研究的主题较为集中，选题的可持续研究强；低被引和低下载文献大多数为专业水平相对较低的作者所创作，论文之间较少有重复的关键词出现，研究主题较为分散，论文的选题较独特，但进行后续研究的不多。下载和引用的科研群体特征还有许多可以挖掘的地方，仍需后续研究不断努力深入。相关内容本书在第一章做了重点论述。

学术期刊排名和论文评价被广泛应用于不同的需要。科研用户可以使用它们寻找最适合的期刊来发表他们的论文,与此同时,科系、学院或者大学越来越多地在招聘、科研评估、评审奖励、项目资金分配等过程中使用论文或期刊的评价指标,包括最常见的影响因子 (impact factor, IF)、相对引用率 (relative citation ratio, RCR)、期刊被引频次 (cited number of journal)、期刊即时反应速率 (immediate reaction rate of journals)、平均引文率 (mean citing rate)、期刊他引率 (non-self-cited rate)、期刊被引半衰期 (cited half-life) 等。相关内容本书在第二章做了论述。

全媒信息环境下科研用户充分享受了获取信息的自由和个性,充分利用了自己的碎片时间。然而结合我们的调查可以发现,全媒信息环境并没有让科研工作者感觉轻松,反而让他们逐渐失去掌握信息的主动权。如他们对手机的依赖程度越来越高,独立而不受干扰的时间越来越少。如何提高科研用户的信息素质和素养,让他们在全媒信息环境下如鱼得水,在本书的第三章做了探讨;而第四章重点探讨了科研工作中学术不端的问题,这是值得深思的。第五章主要探讨与高校科研工作密切相关的图书馆信息服务,包括 ESI 和 InCites 数据库的使用等,以期为更多的科研用户提供思路。

本书是作者结合多年信息服务工作岗位进行研究的总结和回顾。

本书的出版得到了如下科研基金资助:

* 广东工业大学青年重点基金项目“基于高校科技创新及学科建设的信息服务模式研究”(项目编号:15QNZD009)

*国家自然科学基金项目“单元化流水车间双重资源优化配置”(项目编号:71402033)

*广东省教科文卫工会项目“阅读疗法对高校女职工情绪亚健康的研究”(项目编号:2018jkwgh029)

*广州市哲学社会科学发展“十三五”规划2018年度课题“基于阅读疗法的职场女性情绪亚健康的研究——以广州高校女职工为研究样本”(项目编号:2018GZGJ46)

目录

第一章 科研用户下载和引用论文的研究	1
第一节 论文标题与科研用户下载和引用的关系	1
第二节 论文下载和引用的科研用户群体特征研究	9
第三节 科研论文下载与阅读情况调查分析	21
第四节 图书被引用特征探析	31
第二章 论文和科研用户关系的研究	39
第一节 国内有关论文和科研用户关系的研究	39
第二节 国内有关论文和科研用户关系研究的部分观点	46
第三节 国外有关论文和科研用户关系的研究	59
第四节 国外有关论文和科研用户关系研究的部分观点	64
第三章 搜索引擎与用户的关系	69
第一节 搜索引擎对国内网络信息传播的影响	69
第二节 竞价排名与用户体验的关系	75
第三节 掌媒环境下信息素质教育的新内容	82
第四节 成教学生信息素质调查分析	89

第四章 论文与科研工作者学术不端问题研究	95
第一节 论文与科研工作者学术不端问题	95
第二节 学术不端问题研究的观点	103
第三节 论文学术不端署名问题探析	107
第四节 理性利益与师德建设	112
第五章 高校科研用户信息服务	118
第一节 查收查引与论文分区和影响因子	118
第二节 科技查新服务	128
第三节 ESI 和 InCites 数据库的使用	133
第四节 图书馆为科研用户服务的核心竞争力探析	142
第五节 传播学模式下高校图书馆服务的其他功能探析	147
附录	153
附录 1:《SCI 论文收录及他引次数检索证明》	153
附录 2:《SCI 论文收录及分区和影响因子检索》	156
附录 3:ESI 分析结果广东工业大学高被引论文前 20 列表	158
参考文献	163

第一章 科研用户下载和引用论文的研究

第一节 论文标题与科研用户下载和引用的关系

【本节要点】本小节结合英文论文标题的已有研究成果，深入探讨中文论文标题，通过数据统计定量分析中文论文标题类型的下载和引用的特点，以及标题长度和带冒号的标题对下载和引用的影响，揭示论文标题与下载和引用的关系，填补国内关于这一研究的空白。

一、关于论文标题的研究

2011年8月的外文期刊《科学计量学》(Scientometrics)发表了Jamali和Nikzad合作的研究论文 *Article title type and its relation with the number of downloads and citations*^①。他们抽取了2007年美国科学公共图书馆(the Public Library of Science, 以下简称 PLoS)六份期刊的2172篇论文进行分析和研究,把论文题目类型与下载量和引用量同时进行比较,进而发现:不同题目类型(结论型、中立型和问题型)的科技论文在下载量和引用率方面存在显著差异,其中问题型题目的科技论文下载量更多但引用率却相对较少;题目较长的科技论文的下下载量稍少于题目较短的科技论文;带冒号的科技论文的题目更长,其下载量和引用率也相对较少;题目的特征对论文下载量的影响强于引用量。

尽管国外关于文章标题的研究已从不同的角度展开,包括语法、科学计量和学术交流等,但已有研究大多数是从单个方面对论文标题的影响进行分析,很少有如Jamali和Nikzad的研究同时考察论文标题对其下载和引用的影响,国内尚没有关于这方面的定量研究,甚至关于论文标题的研究目前在国内也很少见。此外,PLoS六份期刊收录的是英文论文,内容分

^① Hamid R Jamali, Mahsa Nikzad. Article title type and its relation with the number of downloads and citations [J]. Scientometrics, 2011, 88(2): 653-661.

别来自 PLoS One、PLoS Biology、PLoS Medicine、PLoS Computational Biology、PLoS Genetics、PLoS Pathogens, 研究结果是针对英文文章而言的, 中文的相关结论是什么呢? 或者说中文科技论文的标题是否也体现了 Jamali 与 Nikzad 的研究结果呢? 中文与英文之间有没有差异呢? 带着这些问题, 笔者同样选取 2007 年 CNKI 收录的 6 个与生物学相关的领域 (生物学工程、遗传学、细胞学、生物化学检验、临床检验、仿生学) 进行了统计研究, 以期发现其中的异同。

为限定研究对象为论文标题, 笔者逐条阅读和辨认相关数据, 剔除了所有的社论、勘误、信件、政策、评论和新闻等, 从而获取相关论文标题 3576 个。本研究将在此样本基础上探讨中文论文标题类型, 以及标题长度和带冒号的标题对下载和引用的影响。

二、标题类型与下载和引用的关系

尽管中英文语境不同, 但中文论文的标题类型与英文一样也可划分为三类: ①结论型标题 (declarative title) 不只是涵盖研究主题, 还将论文想要阐述的主要结论进行说明, 如“变量 X 高度正相关影响变量 Y ”; ②中立型或描述型标题 (descriptive or neutral title) 只对论文的主题进行介绍, 并不涉及论文结论或研究结果, 如“基于……的研究”之类的; ③问题型标题 (question title) 采用疑问式来指示论文的主题以吸引读者的注意, 如“什么决定了变量 Y 的高度? ”。在统计的 3576 个标题里, 经过仔细辨认而确定有中立型标题 3203 个, 结论型标题 368 个, 问题型标题 5 个。

为统计标题类型与下载和引用的关系, 笔者选用平均数作为统计值, 在统计中发现有些论文的下载次数或引用次数特别高, 有极端值出现, 这可能会把平均数的代表性削弱。为了排除这一干扰, 本研究利用散点图将这些数据剔除, 如图 1-1-1 和图 1-1-2 所示, 脱离群体点较远的点将被剔除。经过筛选以后, 保留了下载次数低于 1512 以及引用次数低于 40 的文章标题。筛选后中立型标题减少了 8 个, 结论型和问题型标题保持不变。可见在这一统计领域里, 极端高引用和高下载的文章都是中立型标题。

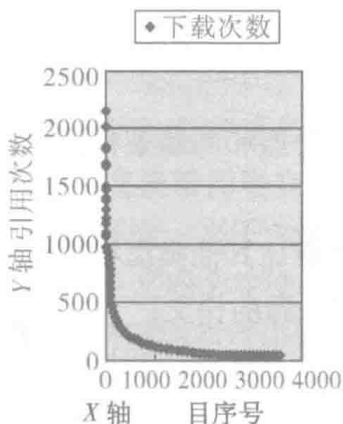


图 1-1-1 下载次数散点图

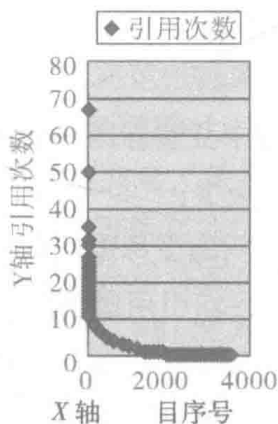


图 1-1-2 引用次数散点图

经过整理和统计，中文标题类型的下载和引用结果统计如表1-1-1所示。为了便于比较，笔者将 Jamali 和 Nikzad 在英文标题类型的下载和引用结果整理成表 1-1-2。

表 1-1-1 中文标题类型与下载和引用的关系

标题类型	百分比	个数	平均下载次数	平均引用次数
中立型	89.3%	3195	104.8	1.96
结论型	10.3%	368	103.9	1.92
问题型	0.1%	5	152	2.75

表 1-1-2 Jamali 和 Nikzad 的英文标题类型与下载和引用结果统计

标题类型	百分比	个数	平均下载次数	平均引用次数
中立型	67.16%	1442	3906	16.92
结论型	30.74%	660	3588	16.93
问题型	2.1%	45	5817	10.47

通过对比两个表格我们可以发现,中文和英文标题类型有如下几个异同点:

(1) 两者中立型标题占较高比例,结论型和问题型标题较少,而中文的问题型标题极少;

(2) 两者问题型标题的科技论文下载量比其他类型标题更多,而中文问题型标题的平均引用量也高于其他类型标题的论文;

(3) 两者中立型标题的平均下载量稍微高于结论型标题,而中文的中立型的平均引用量也稍高于结论型标题。

结合两个表格对比的异同,笔者尝试分析原因:

首先,中立型标题仍然是中英文语境学者和编辑较感兴趣的。这类标题只需对研究对象、论文的主题或中心论点进行介绍,适用范围非常广泛,题目的开放性和收容性比较强,不受研究层次和研究深度的限制,只需抓住主要研究对象进行描述即可,字数可长可短,不需要将结论或成果展现在标题中。既体现了研究内容的可拓展性,又可以给论文作者或读者保留探讨的空间,正如马克思主义哲学所认为的“宇宙中的各个具体事物和每个具体过程都是有条件的、有限的、相对的,而整个宇宙的存在和发展又是无条件的、无限的、绝对的”^①。也即任何一个作者都会避免鲁莽地给自己的研究下结论,因为受到研究条件和时间等因素的限制,特别是在标题字数有限的条件下更应慎重,从而避免直接受到反驳的可能。其次,并非每篇论文都一定有结论或成果出现,所以结论型标题的使用是受到局限的。某些专业领域(医药或公共卫生等)的编辑对于接受结论型的标题可能会更加谨慎,因为可能对公众健康产生影响。此外在学术论文标题中使用问号一般是不被建议的,所以,问题型和结论型标题较少,中立型标题占较高比例。且由于中国人的思想表达较为含蓄,结论型标题在中文领域所占的比例会更低。

国外另一学者 Ball 在 2009 年研究了 1966 年到 2005 年的近 20 万篇 SCOPUS 文章的物理学、生命科学和医学科学的标题,发现带问号标题的

^① 赵凤岐,按苗.论绝对与相对[M].南京:江苏人民出版社,1982.

文章在过去的 40 年数量显著增加,其增长率从 50% 到超过 200%^①。Ball 认为问号在科学文章标题中被不断使用,背后的决定性因素之一是对文章的营销和推广。中文标题问号的增长率笔者并未作统计,但是从获取的 3576 篇文章中,直接带问号在标题上的仅有 2 篇文章,其余 3 篇都是省略了问号的问题型标题。至于为何省略了问号,这可能与我国科技论文编辑要求有关,学术论文题名中使用标点符号“不符合科技期刊编排标准对题名的一般要求,是题名写作中的大忌”^②。甚至在 CNKI 或其他数据库检索时,我们也可以发现在关键词或篇名等检索中是不支持问号的。因为在检索语言里,问号是用来表达截词符或其他相关的检索算符,而不是中文的标题内容的提问。英文标题一般不建议使用问号^③,但并不违反科技期刊编排标准,而在中文标题中是违反了学术期刊编排标准且是大忌。此外,在我们的思维里,问号可能被视为不严谨或研究还未有定论,也无成果,而科研论文是学术性和严谨性或成果性的代表,并不适合以提问的形式表达,所以中文的问题型标题特别少。

论文的作者在标题中使用提问或质疑可以引起读者的兴趣或好奇心,激发他们阅读的欲望,进而获得更多的下载量,但并不是每个下载都意味着引用,甚至下载不一定意味着使用或阅读文章。笔者曾持续跟踪调查有关高校科研工作者,统计出有 90.2% 的人不会全部阅读使用下载的文章^④。在科研实践中,引用只有文章作者才做,但引用实际上意味着这篇文章与作者写的东西主题相关,它已经通过他的质量控制,甚至可以佐证他的观点。因此,问题型标题不足以吸引别人马上引用,但它可能吸引别人下载,故问题型的标题论文下载量都较高。但中文的问题型标题为何下载和引用都偏高呢?笔者下载了 5 篇文章,仔细阅读发现,5 篇文章均是介绍或推广某项国外新技术研究或新成果,而这些技术对于解决某些关

① Ball R. Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966 - 2005 [J]. *Scientometrics*, 2009, 79(3): 667-679.

② 蔡建霞,刘荷芬,王玉霞.科技论文 名写作与投稿[J].*地域研究与开发*,2008(01):126-128.

③ Gustavii B. How to write and illustrate scientific papers [M]. Cambridge: Cambridge University Press, 2008: 2.

④ 林佳瑜.论文下载次数与阅读使用次数的调查分析[J].*图书馆杂志*,2012(03):36-39.

性问题是 有较 大帮助 的, 由于其实用性和新颖性以及重要性从而获得国内学者的下载和引用, 所以中文问题型标题引用量高于结论型和中立型标题。此外还发现问题型标题均出现在实用型刊物中, 基础理论研究的刊物并未见, 这也反映了国内基础理论研究的刊物可能不会乐于接受问题型标题。但由于笔者此次的统计里仅有 5 篇问题型标题, 其代表性可能不足以说明其他领域的情况。

其次, 中立型标题因其观点含蓄, 容易引起读者的注意, 从有歧义的标题去寻找某个确定的点可以吸引用户好奇心。当用户检索到标题, 考虑潜在有关他们的信息需求的文章, 但无法从标题或摘要获得足够的信息, 他们可能会决定下载文章和浏览它, 所以可能导致更多的下载次数。结论型的标题通常有一个明确的发现, 从标题就可以明确获知其研究结果。如果标题是清晰的、明确的, 并已有结论展示, 那么用户可以直接得到他们所需要的, 不需要阅读全文, 从而导致结论型的标题下载较少。这也可以解释中立型标题的平均下载为何稍微高于结论型标题。由于中立型标题的下载高于结论型标题, 从而会获得更多阅读者和使用者, 从总体趋势上来说, 高下载会带来高引用, 因此中立型标题的引用量也会比结论型标题高。

三、标题长度与下载和引用的关系

为统计标题长度对下载和引用的影响, 笔者先把所有标题的引用次数和下载次数排序, 并把两者排前 40 位和排最后 40 位的标题找出, 并进行相关统计。统计得出, 无论是按引用或下载排序, 前 40 位的标题字数都少于后 40 位的, 也即前 40 位的标题较短, 后 40 位的标题较长, 由此可以推出题目较短的科技论文的引用次数和下载次数比题目较长的科技论文多。这与 Jamali 和 Nikzad 研究出来的结果是一致的, 也即题目较长的科技论文的下载量稍微少于题目较短的科技论文。在中文标题里, 不仅较短的科技论文会获得较高的下载量, 还能获得较多的引用次数。这是什么原因呢?

笔者逐条研读下载和引用排前 40 位和后 40 位的标题, 发现前 40 位的标题有不少是文献综述, 后 40 位的标题没有发现文献综述。统计结果如表 1-1-3 所示, 下载排前 40 位的标题有综述的 28 个, 引用排前 40 位有综述的 15 个, 综述标题的平均字数比前 40 条和后 40 条的标题平均字数要少。综述条数越多, 标题的平均字数就越少, 进而标题的平均长度就

越短。由此,较短题目的科技论文下载次数和引用次数较多,较长题目的科技论文下载次数和引用次数较少。综述为何能获得较高的下载量和引用量呢?笔者认为综述是综合起来介绍某领域的研究概况,所以题目比其他科技论文的题目较短。另外,综述是反映某一课题的新水平、新动态或新发现,从研究的历史到现状乃至目前存在的问题都要进行全面的介绍和评论,可以让其他相关课题的研究者获知自己的选题在当前研究中的位置及优势甚至突破点等,很多研究者可以从中寻找自己课题的相关支撑。由于综述具有这些特点,所以能获得较多的下载量和引用量。

表 1-1-3 标题长度与综述统计

标题	平均标题字数	字数(不计空格)	综述	综述标题平均字数
引用前 40 名	19	759	15	15
引用后 40 名	24	957	0	—
下载前 40 名	15	601	28	14
下载后 40 名	23	908	0	—

四、有冒号的标题与下载和引用的关系

中文论文带有冒号的标题并没有英文多。在 3576 个标题中,笔者仅发现了 26 个带冒号的标题,仅占 0.73%,而 Jamali 和 Nikzad 的统计中有 14.96% 的标题带冒号。经过统计,笔者得到表 1-1-4,有冒号的标题平均字数比没冒号的标题多 7 个字,而其平均引用次数和下载次数都高于没冒号的标题。即在统计的中文标题里带冒号的标题虽然更长,但平均下载次数和平均引用次数也更高。这与题目较长的科技论文的下载量和引用量稍微少于题目较短的科技论文是矛盾的,具体原因又是为何呢?笔者将论文的作者进行统计,发现有冒号的标题来自核心刊的有 88.5%,论文有合作者的占 80.76%,没冒号的标题来自核心刊的占 65.6%,论文有合作者的占 71.73%。由此我们可以判断,有冒号的标题论文大多数是由研究团队共同完成,而且大多数发表在核心刊,其研究深度和广度可能大,进而获得更多的下载和引用。

表 1-1-4
 有冒号和没冒号的标题的下载和引用

类型	个数	平均标题字数	平均引用次数	平均下载次数	合作者	核心刊
有冒号	26	27	3.38	162.96	80.76%	92.3%
没冒号	20	20	1.93	104.91	71.73%	65.6%

五、 结语

通过调查样本的统计，我们发现中文论文标题与英文论文标题对下载和引用产生很多相似的影响，比如不同题目类型的科技论文在下载量和引用量方面存在显著差异，其中问题型题目的科技论文下载量更多但引用量却相对较少；题目的特征对论文下载量的影响强于引用量；中立型标题占较高比例，结论型和问题型标题较少；中立型标题的平均下载量稍微高于结论型标题。此外，在中文论文标题的统计中出现了如下与英文论文标题差异的地方：中文标题的问题型标题极少；中文问题型标题的平均引用量高于其他类型标题的论文；中文的中立型标题平均引用量稍高于结论型标题。且从平均数的统计结果来看，高下载的文章趋向高引用。

标题是用户对信息做出相关判断和检索的主要来源，标题的质量会影响研究性文章的效果，甚至直接影响下载和引用。关于标题的研究目前国内还是空白。尽管有不少文章都提及标题的重要性，甚至在我们的观念里都认为标题很重要，但如何体现重要，如何用数据进行定量分析，这在国内还是未见的。中文和英文标题类型的相同特点展示了学术研究尽管存在语境的区别，但是人类对知识和真理的追求是一致的，对待科学研究的严谨态度是没有区别的。而差异的出现可以带给我们思考，甚至可以帮助我们探索差异产生的深层原因，以帮助我们更好地进行学术交流。关于论文标题的研究还有许多可以深入的地方，比如对论文标题的语法和内容分析^①，或者对作者的态度进行深入的定量研究，等等。这些都可以帮助我们揭示很多群体的特征，需要在后续研究中不断完善。

^① Jacques T S, Sebire N J. The impact of article titles on citation hits: An analysis of general and specialist medical journals[J] . Journal of the Royal Society of Medicine Short Reports, 2009, 1(2) : 1-5.