

- effects on peanut allergenicity [J]. Food Chemistry, 2012, 132: 360-366.
- [32] 丛艳君. 花生主要致敏原致敏机理及其脱敏方法的研究 [D]. 北京: 中国农业大学, 2007.
- [33] Jianmei Y, Mohamed A, Ipek G. Enzymatic treatment of peanut kernels to reduce allergen levels [J]. Food Chemistry, 2011, 127 (3): 1014-1022.
- [34] 胡纯秋, 高金燕, 罗春萍, 等. 重组花生致敏原 Ara h2 的生物合成 [J]. 食品科学, 2010, 31(15): 203-207.
- [35] 易海涛, 刘志刚, 刘芳, 等. 花生致敏原 Ara h2. 02 的克隆、表达及免疫学鉴定 [J]. 江西师范大学学报, 2010, 34(5): 531-534.

## 综述

# 食品安全风险监测数据统计处理常见问题探讨

宫春波, 王朝霞, 孙月琳, 董峰光

(烟台市疾病预防控制中心, 山东 烟台 264003)

**摘要:** 为了保证食品安全风险监测数据处理、分析的合理性和科学性, 获得可靠的结果, 对低于检出限数据的处理、率的比较、样本类别的分组所适用的统计分析方法进行探讨。总结了食品安全风险监测数据的性质及特点, 阐明了按特定研究目的数据处理分析所适用的统计分析方法, 按特定的分析目的选择合适的统计分析方法, 从而获得可信、可靠和可比的结果, 为监测提供有力的数据支持。避免盲目的运用统计分析方法评价分析食品安全风险监测数据, 进而导致评价结果偏差或错误, 误导政府决策、部门监管和消费者消费。

**关键词:** 食品安全; 风险监测; 数据; 统计

**中图分类号:** R195.1      **文献标识码:** A      **文章编号:** 1004-8456(2013)06-0575-04

## Application of statistic analysis processing on food safety risk surveillance data

GONG Chun-bo, WANG Zhao-xia, SUN Yue-lin, DONG Feng-guang

(Yantai Center for Diseases Control and Prevention, Shandong Yantai 264003, China)

**Abstract:** This article aims to promote the rational and scientific statistical analysis applied to food safety risk surveillance. The not-detected data, rate comparison, sample category and statistical analysis methods are discussed, and the characteristics of food safety risk surveillance data are analyzed and summarized. It is very important to apply a suitable and correct statistic analysis to obtain reliable, credible and comparable results, which would provide powerful support for monitoring. Otherwise, the conclusion of food safety risk surveillance would mislead the government supervision decisions, and consumers' choices.

**Key words:** Food safety; risk surveillance; data; statistics

食品安全风险监测主要对食源性疾病、食品污染以及食品中的有害因素进行监测<sup>[1]</sup>, 进而进行风险评估, 为制定、修订食品安全标准和对食品安全实施监督管理提供科学依据。随着《食品安全法》的颁布和实施, 食品安全风险监测工作在全国范围内全面开展, 监测数据的统计分析和结果报道陆续见于期刊。但对于低于检出限数据的处理、率的比较、均值的计算以及样本归类等处理, 未见通用的、合适的处理方法, 监测数据的统计处理方法差异较大, 进而存在分析结果的偏移和结论的偏差。因此, 为了获得较为准确的食品安全风险监测分析

结果, 为风险评估和风险预警提供可靠的数据分析, 利于地域间食品安全水平的比较和污染原因的分析, 食品安全风险监测数据的统计分析和处理需要规范化, 以保证分析评价结果的可靠性和可比性。

### 1 低于检出限数据的统计处理

食品安全风险监测中涉及到的化学污染物和有害元素的监测指标, 常利用痕量和超痕量分析法。为评价和比较各种分析方法, 表征检测能力, 在同一分析系统中提出了两个不同的性能参数<sup>[2]</sup>: 检出限 (LOD) 和定量限 (LOQ)。限于分析方法的检测能力, 样本检测过程中必然出现低于检出限的检测值。目前, 对低于检出限的数据处理, 常见的方法为“零”代替, 并确定为“未检出”样本, 由于被

测样本中监测指标的分布介于真“零”与 LOD 之间,所以“零”值代替“未检出”个数进行数据分析不甚合理,特别是偏态分布的数据,“零”值代替“未检出”个数无法实现对数转换计算<sup>[2]</sup>。为获得可靠的统计和分析结果,WHO 全球环境监测系统/食品污染监测与评估规划(GEMS/FOOD)第二次会议上提出了“食品中低水平污染物可信评价推荐”<sup>[3]</sup>,这是目前常用的食品检测中低于检出限数值的处理参照方法,应用中也多为“零”或“1/2 LOD”替代低于检出限数值,进行分析和统计。

### 1.1 检测限值应基于相同或相近的方法

目前,食品安全风险监测由国家层面统一部署,各级主管部门逐级开展。检验方法的选择遵循“优先选择最灵敏的国家标准检测方法”的原则,也可以选择比国标更准确更灵敏的非标准方法,并且使用前需要进行验证确认<sup>[4]</sup>。但是,由于地域经济发展程度、各参与实验室的投入和人员配备均不同,检测结果必然存在差异。加之,食品安全风险监测的主要任务是搜集区域性、群体性、基础性的数据和相关信息,为有关部门提供决策依据和技术咨询,为风险评估与标准制修订提供科学数据<sup>[5]</sup>。笔者认为,选用最灵敏的国家标准检测方法开展食品安全风险监测,能够满足其任务和目的的需求。因此,建议规范、统一各参与实验室检测方法,以工作手册推荐方法为主,力求参与食品安全风险监测实验室的检出限值均为基于国家标准方法规定的数值,或者更优于国标方法的检出限值,介于 LOD 和 LOQ 之间结果均以数值报告。以期保证参与的实验室采用相同的或相近的检测方法,获得基于同

样的检测限和测定限的检测结果,为进一步的数据分析统计和结果的比较,提供一个相同的参照数值。

### 1.2 选择合适的统计方法

目前,处理低于检出限数据的方法多为简单替代法,即以“零”、“1/2 检出限”或“检出限值”替代<sup>[2,6]</sup>。利用简单替代法估算数据组均值、标准偏差,往往会引入较大偏差或导致错误结果<sup>[6]</sup>,进而误判污染水平和污染趋势,不能正确地评价其风险,达不到监测的目的。同时,也不能真实反映地域差异导致的污染程度,影响污染来源、途径等的分析和确认。因此,食品安全风险监测数据的处理应根据监测数据组特点,选择合适的数据处理方法。例如,百分位数法、参数估计法、拟合优度法、分布法<sup>[6]</sup>,低于检出限数值的处理方法及其要求见表 1。其中,均值配合标准差以及中位数法配合四分位数间距法的统计分析,适合低于检出限结果 25% 的数据组,能够较好地反映数据的集散或离散趋势,计算简单,表征性好,契合食品安全风险监测数据组的特征,值得关注和使用。但是,当低于检出限数据高于 50% 时,中位数和四分位数间距不可求,应用上受到数据条件的一定限制<sup>[6]</sup>。

通过分析烟台市 3 年的食品安全风险监测结果数据和参考相关文献,发现化学污染物及有害因素的监测数据中,低于检出限的数据比例范围为 0% ~ 100%,无关样本种类、检测指标。因此,食品安全风险监测数据处理过程中,要优先考虑低于 LOD 数据所占的比例,参考相关文献或权威部门推荐的处理方法,进行数据的统计分析,确保结果的可靠性、可信性和可比性。

表 1 不同比例的低于检出限监测数据的处理方法

Table 1 The statistic analysis processing method for different datas belowed detection limits

结果 < LOD 的比例 (Z)	适合的处理方法	文献来源
Z = 0	均值、中位数、标准差、最大值、最小值	[2]
0 < Z ≤ 25%	< LOD 的结果按“1/2 LOD 值”计,并标注;均值、标准差、中位数或四分位数间距;最大值,最小值 (< LOD 值)	[2,7]
25% < Z ≤ 50%	< LOD 的结果按“1/2 LOD 值”计;截尾均值法;极值推压法;Aitchison's 方法(预报拟合优度准则); log-probit regression method	[2,6,8]
50% < Z ≤ 80% <sup>a</sup>	< LOD 的结果按“0 ~ LOD 值之间任何值”计;对数正态分布法、边界值分析法	[2,6,8]
Z > 80% <sup>b</sup>	< LOD 的结果按“1/2 LOD 值”计,并标注;均值	[8]

注:a 表示数据组至少 25 个数值结果,样品数 < 100 时,慎重处理;b 表示没有合适的统计处理,能获得 95% 置信区间

## 2 食品类别

食品安全风险监测结果数据统计分析时,均涉及计数和计量资料的统计指标。样本按照“统一的分类依据”进行合理的分组,是计数和计量资料数据进行统计分析的前提,加之率的标准化条件之一是已知各类食品的合格率,也体现了样本类别分组的重要性。

目前,食品的分类标准、分类方法以及分类依据差异性较大,同一食品往往可以作不同类的分组。而不同食品种类的污染物标准限量存在差异,因此不同的食品分类方法将影响监测结果,不利于样本检测数据的统计分析。鉴于食品安全风险监测实施样本归类网络直报的现状,建议样本统计分组以“全国食品污染物监测数据汇总系统平台

(26 大类,100 亚类,约 1 288 小类)”和“全国食源性致病菌监测数据汇总系统平台(17 大类,93 亚类,约 282 小类)”的食品类别为基础,剔除重复或按地域分类的类别,加以整合、统一归类并应用到数据统计分析中。笔者认为样本监测数据统计分析时,按照亚类或者小类进行分组分析,能够更好地反映监测结果。统计分析分类时,遵循“大样品数,按照小类分;小样品数按照亚类分”的原则。例如:样本的分组在满足统计方法的条件和要求下,肉与肉制品监测结果数据分析时,若样品数较大,按照猪肉、牛肉、羊肉、鸭肉、鸡肉等小类分析,会获得较好地结果;由于目前监测样本相对较少(特别是地市级),按照生禽肉、生畜肉、熟肉制品 3 类分组分析,也能够较好地判定其安全质量结果,利于评价和比较其安全质量水平。

### 3 统计分析方法的正确应用

#### 3.1 统计分析方法的正确选择

基于分析目的不同,食品安全风险监测数据资料存在着计量和计数资料的转换。例如,分析肉及肉制品和水产品中铅污染的比较时,其样本铅含量数据为计量资料;而按照国家标准铅的限量,判定其合格数后,合格率的数据资料则为计数资料。数据资料的性质差异,决定了其统计方法的不同运用。

##### 3.1.1 计量资料的统计分析

计量资料的统计指标分为描述集中趋势和描述离散趋势的指标,前者包括平均值、中位数、众数等,后者主要是标准差、方差、全距、四分位数间距、极差等<sup>[9]</sup>。当然,各个指标具有独自的应用条件,应当正确区分和应用,可以将两类统计指标有机地配合使用,实现全面地描述一组数据的分布特征<sup>[10]</sup>,通常均数可与标准差、极差、变异系数等搭配使用;中位数可与四分位数间距、极差等搭配使用等<sup>[9]</sup>。计量资料符合正态分布或者均匀分布时,用“均数±标准差”;而偏态分布的计量资料则采用中位数、四分位数间距和极差等加以描述;不能确定数据的分布类型时,应选用中位数、四分位数间距进行统计描述<sup>[11]</sup>。参数检验法和非参数检验法是计量资料常用的统计学方法,前者主要为  $t$  检验和  $F$  检验,后者主要为秩和检验。两组间样本均数的比较常用  $t$  检验和  $u$  检验,而两组以上样本均数比较时,用方差分析<sup>[11-12]</sup>。

##### 3.1.2 计数资料的统计分析

绝对数和相对数是描述计数资料的统计指标。绝对数为原始资料经汇总得到的小计或总计数,不能深入地分析现象间的关系和变化发展情况。相

对数则是两个有关的绝对数之比,表示相对大小,主要包括率、构成比和相对比。 $\chi^2$  检验常用于分类计数资料的假设检验,即用于 2 个样本率、多个样本率、样本内部率的比较。常用的  $\chi^2$  检验可以分为 3 类:2×2 表  $\chi^2$  检验、配对资料  $\chi^2$  检验和 R×C 表  $\chi^2$  检验<sup>[11-13]</sup>,其中 2×2 表  $\chi^2$  检验和 R×C 表  $\chi^2$  检验适用于食品安全风险监测数据的统计分析。

#### 3.2 率的标准化的应用

标准化率(standardized rate)又称标化率或调整率(adjusted rate)。用于两个及两个以上率或构成比比较,在其他非研究因素各组间有不同,且非研究因素会干扰研究因素的分析时,可用标准化率进行比较,消除非研究因素的影响,得到反映真实情况的结果<sup>[14]</sup>。食品安全风险监测是持续的、动态的监测,探求食品污染的趋势和规律是其目的之一,必然进行率相关的比较,以期掌握污染水平、污染规律和趋势。表 2 是某市 3 年来的化学污染物及有害因素监测结果统计,2010—2012 年的粗合格率分别为 94.21%、87.09% 和 91.79%,其污染趋势为“Λ”形,原因在于样品数量、样本类别等不同所致。而对各年度的率标准化后,得到的合格率分别为 94.21%、86.44% 和 85.73%,则直观表明合格率逐年下降。食品安全风险监测结果中,两个或两个以上相关率的比较时,倡导将标准化率应用到食品安全风险监测结果的率比较中,以期分析统计时尽可能消除各样本间存在的差异,真实反映监测结果的变化趋势,获得可信的结果,为食品安全风险监测提供可靠的数据支持。

表 2 结果表明,率的标准化可以消除年度间样本种类、样本构成比的差异,获得准确的年度调整率,其结果更合理。率的标准化应用条件为已知各类样品的合格率,标准组选择合适。对于后者,可以选择全国、全省年度样本类别、任务分配数量为标准组,也可以选择将比较的样本类别的数据合并为标准组。常用的率的标准化法为直接法和间接法,当标准为数据或数据构成时,用直接法计算标准化率;当标准为率的指标时,用间接法求得标准化率。

### 4 小结

食品安全风险监测作为一项国家制度,在全国范围内是逐年拓展趋于完善,监测样品数逐年增加,监测数据也随之增多。随着风险评估、风险交流和风险预警的深入开展,需要可靠的、可信的监测结果来支撑其判定和结论。因此,科学地处理、分析食品安全风险监测数据,获得可信的评价结果尤为重要。化学污染物和有害毒素监测数据的统

表2 某市2010—2012年食品监测粗合格率与调整合格率的比较

Table 2 The results of food safety risk surveillance datas after rate standardization of a city during 2010-2012

类别	标准组	构成比	2010				2011				2012			
			样品/份	粗合格率/%	调整合格率/%	期望合格数 <sup>a</sup>	样品/份	粗合格率/%	调整合格率/%	期望合格数 <sup>a</sup>	样品/份	粗合格率/%	调整合格率/%	期望合格数 <sup>a</sup>
粮食	310	0.128	30	100	12.82	310	150	98.67	12.65	306	130	98.46	12.62	305
水产品	366	0.151	101	99.01	14.99	362	107	85.05	12.87	311	158	81.01	12.26	297
肉类	510	0.211	116	88.79	18.73	453	133	65.41	13.80	334	261	93.87	19.80	479
乳制品	111	0.046	10	100	4.59	111	18	100	4.59	111	83	86.75	3.98	96
蔬菜	369	0.153	103	88.35	13.48	326	74	94.59	14.44	349	192	95.83	14.62	354
熟制坚果	35	0.014	10	80	1.16	28	10	100	1.45	35	15	100	1.45	35
酱腌蔬菜	22	0.009	10	100	0.91	22	12	100	0.91	22	0	0	0.00	0
蜜饯	57	0.024	7	100	2.36	57	20	95.50	2.24	54	30	100	2.36	57
酒类	36	0.015	16	100	1.49	36	20	100	1.49	36	0	0	0.00	0
调味品	114	0.047	15	100	4.71	114	27	100	4.71	114	72	100	4.71	114
婴幼儿类	120	0.050	13	100	4.96	120	25	100	4.96	120	82	96.34	4.78	116
豆制品	69	0.029	22	100	2.85	69	37	32.43	0.93	22	10	100	2.85	69
米面食品	118	0.049	13	84.62	4.13	100	25	80	3.90	94	80	71.25	3.48	84
饮料	35	0.014	13	100	1.45	35	12	100	1.45	35	10	100	1.45	35
餐饮食品	113	0.047	26	100	4.67	113	87	100	4.67	113	0	0	0.00	0
冷冻饮品	33	0.014	13	100	1.36	33	10	100	1.36	33	10	100	1.36	33
合计	2 418	1	518	94.21	94.67	2 289 <sup>b</sup>	767	87.09	86.44	2 090 <sup>b</sup>	1 133	91.79	85.73	2 073 <sup>b</sup>

注:a表示直接标准化法;b表示标准化后各期望合格数之和

计分析多为计量资料的统计分析,特别是低于检出限数据的处理,应该分析数据组中低于检出限数据的所占比例,选择合适的指标(平均指标或变异指标)加以描述,运用合适的统计方法比较和评价其结果;尤其是低于检出限数据所占比例较大,在描述指标选择以及结果判定时尤为慎重。食源性致病菌监测数据往往为计数资料,相对数的描述指标比较多,相关率的比较时应该注意统计方法的选择和运用,选择合适的统计方法。尤其是进行 $\chi^2$ 检验和率的标准化时,要根据数据组的性质和研究目的,选择适用的统计分析方法。切忌盲目地运用统计学方法分析评价食品安全风险监测数据,导致结果的偏差或者错误,误导主管部门监管和消费者消费,避免风险评估和风险预警的偏移。

## 参考文献

- [1] 中华人民共和国主席令(第九号). 中华人民共和国食品安全法[Z]. 2009-02-28.
- [2] 王绪卿, 吴永宁, 陈君石. 食品污染监测低水平数据处理问题[J]. 中华预防医学杂志, 2002, 36(4): 278-279.
- [3] WHO. GEMS/Food-EURO second workshop on reliable evaluation of low-level contamination of food-report of a workshop in the frame of GEMS/Food-EURO. WHO: 1995.

- [4] 杨大进, 李宁. 2013年国家化学污染和有害因素风险工作手册[C]. 北京: 中国质检出版社, 2012.
- [5] 钟凯, 伍竟成, 牛凯龙, 等. 食品安全风险监测与监督抽检相关问题的探讨[J]. 中国食品卫生杂志, 2012, 24(2): 148-151.
- [6] 张文平. 环境监测中低于检出限数据的统计处理方法[J]. 上海环境科学, 1993, 12(11): 38-40.
- [7] CHAO D, NI H G, ZENG H. Parent and halogenated polycyclic aromatic hydrocarbons in rice and implications for human health in China[J]. Environmental Pollution, 2012, 168: 80-86.
- [8] Parikhit Sinha, Michael, Lambert B, et al. Evaluation of statistical methods for left-censored environmental data with nonuniform detection limits[J]. Environmental Toxicology and Chemistry, 2006, 25(9): 2533-2540.
- [9] 胡明, 孙振球, 史静静. 医学科研数据分析与教学中的基本问题[J]. 中华老年医学杂志, 2012, 31(4): 347-350.
- [10] 詹绍康. 率的标准化[J]. 中国实用外科杂志, 2000, 20(2): 122-123.
- [11] 伍亚舟, 张玲, 易东. 医学论文中统计学问题分析与方法的正确选择[J]. 西北医学教育, 2008, 16(1): 161-163.
- [12] 张知洪, 莫建坤, 雷达, 等. 医学论文中常用统计分析方法的合理选择[J]. 检验医学与临床, 2011, 8(11): 1387-1388.
- [13] 潘兴仪. 卡方检验在医学科研中的应用[J]. 广西医学, 2001, 23(6): 1396-1401.
- [14] 邹焰. 本科教学中标准化率的标准选择探讨[J]. 西北医学教育, 2008, 16(5): 117-119.