Biostatistical contributions to gerontological research

A. John Bailer[1,2]

[1]Department of Mathematics & Statistics
[2]Scripps Gerontology Center
Miami University
Oxford, Ohio 45056
T: 513.529.3538
F: 513.529.1493
Email: baileraj@muohio.edu

Seminar: SUMSRI July 1, 2008

Collaborators:

Mathematics & Statistics
Bob Noble, Doug Noe, Mike Hughes, Stefan Stanev [S]

Scripps Gerontology Center and SOC/GTY
Jane Straker, Matt Nelson, Shahla Mehdizadeh, Haesang Jeon [S]

NIOSH
Bob Park, Tim Bushnell

Funding:

Original study funded by the Ohio Dept. of Job and Family Services.

Ohio Long-term care project

State of Ohio: much of this work comes either from administrative programmatic data or projects that they paid to have conducted.

Outline:

0. Background (and "trailers") – range of gerontology problems (NH quality, NH worker injuries, case management of older adults receiving home care)

1. Evaluating resident care needs
2. Study and data structure
3. If gaming, then what is expected?
4. So how can we test this?
5. What did we see?
6. Afterthoughts

0. Background– range of gerontology problems (NH quality, NH worker injuries, case management of older adults receiving home care)

NH quality – Places Rated or Consumer Reports?

NH worker injuries – predicting NH worker injuries … impact of safety equipment

PASSPORT clients – disenrollment? Predicting falls?

NH quality – Places Rated or Consumer Reports?

Bailer, A.J., Straker, J., Noble, R., Hughes, M., See, K. (2007) Considering nursing home quality: places rated or consumer reports? *Chance* **20**: 59-62.

Can nursing home quality be defined using a single scale?

Can quality be defined the same way in all types of nursing homes?

What variables are the important predictors of quality?

What are the relationships among different quality outcomes?

If you are impatient … the answers are: no, not really, many are possible, and distinct dimensions that are not easily captured.

Peer groups of nursing homes? [cluster analysis results]

Nine distinct variables used for clustering including:

* location (rurality)
* ownership type
* percent of Medicaid residents in the facility
* proportion of Medicaid days in county in a particular facility (a measure of market share)
* occupancy rate
* admission rate
* facility size
* resident acuity (a proxy for the amount of skilled care required by residents)
* existence of nursing home competition in the county (measured by the number of nursing home beds/population age 65+ in the county).

Table 0.1: Four Large Clusters of Nursing Homes in Ohio.

Facility Characteristic	Big-Urban-Sick For-profit Nursing Home	Big-Urban-Sick Not-for profit Nursing Home	Non-hospital rehab	Traditional Nursing Home
Location	Urban	Urban	Rural	---
Ownership	For-profit	Not-for-profit	Mostly for-profit	Mostly for-profit
% Medicaid	Low	Low	Low	High
% Medicare	High	High	High	---
Occupancy	---	---	---	---
Admissions	---	---	High	High
Size	Large	Large	Small	Small
Acuity	Fairly High	Fairly High	Fairly High	Low
Beds / 65+ in county	---	---	---	---
# in cluster	398	179	117	85

Quality defined?

Set of quality variables included 18 quality outcomes in the following 5 classes:

* resident satisfaction (reflecting whether a resident would recommend a facility and a general factor for an analysis of a resident satisfaction survey),
* family satisfaction (two analogous variables),
* deficiencies (number in a recent survey and history of deficiencies in the last two surveys),
* clinical indicators (eight variables representing the proportion of residents who are bedfast, had pressure sores, etc.)
* two staffing variables (total direct care staff ratio and registered nurse (RN) ratio).

Factors that appear consistently across NHs are

Resident Satisfaction,

Family Satisfaction,

Deficiencies,

Staffing, and

Clinical Care Indicators.

Despite considering various factor solutions (rotations), no single dominant factor emerged, i.e. no “q-factor” present.

In addition, the important factors also differed somewhat across clusters, although the five factors were reasonable and the description of these factors roughly fell into areas described above.

We believe the data suggest …

nursing home quality is multidimensional, not one-dimensional…

consumers have different preferences for different aspects of care…

→ any comparison of nursing home quality must include multiple factors, rather than a single indicator.

So, we selected more of a *Consumer Reports* model, as opposed to a *Places Rated* model

NH worker injuries – predicting NH worker injuries … impact of safety equipment (Stefan Stanev, Bob Park, Tim Bushnell)

Nursing home workers experience injuries at rates higher than all industries [exception: certain transportation-related industries].

NH worker in Ohio average injury rate: was 6.6 injuries per 100 (nat'l average: 8.9 per 100 workers)

Lost-time injury rate: 1.6 injuries per 100 workers (25^{th} %ile=0.5; 75^{th} %ile=3.2)

Table 0.2: Relative frequency of safety equipment present in nursing homes in Ohio

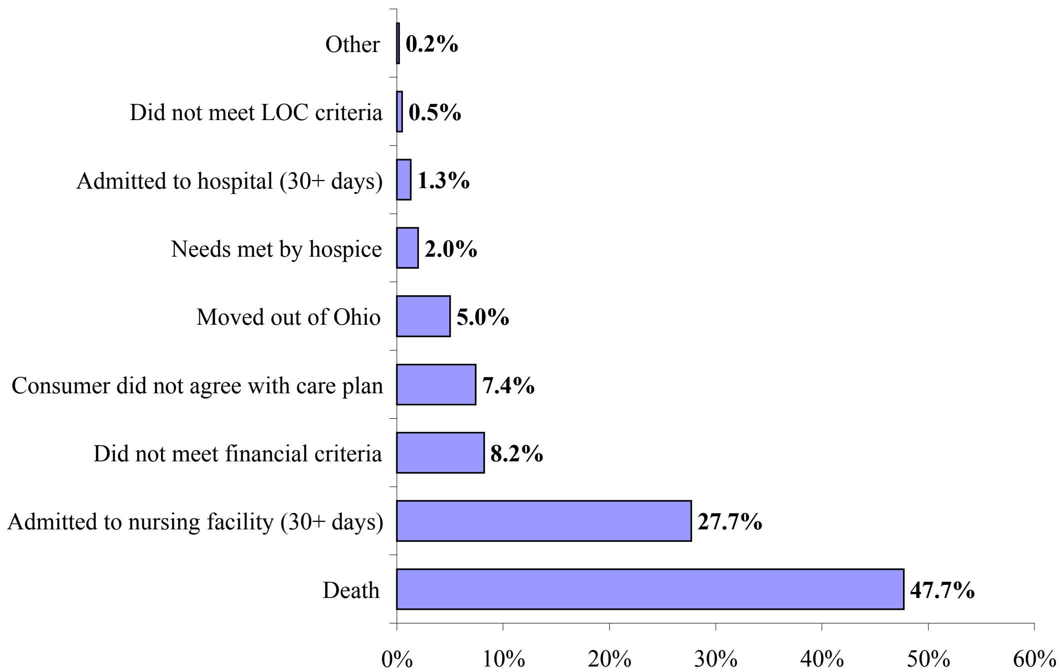
Safety equipment	Percent
Total lift hoist (ceiling mount)	7.3
Mechanical lateral transfer aids	28.5
Friction reducing lateral aids	36.0
Bath lift/easy access bath tubs	54.9
Powered sit-to-stand devices	63.2
Toilet seat adjusted to height of wheelchairs	79.7
Electric Beds	89.6
Total lift hoist (portable)	96.3
"Gait" transfer belts	99.1
109 Missing values	

Table 0.3. Combinations of safety equipment present in Long-Term Care facilities ordered by injury rate per 100 employees.

Equip-ment profile	Total Lift hoist (portable)	Gait' transfer belts	Electric Beds	Toilet seat adj. to height of weelchairs	Friction red. lateral aids	Mech. lateral trasnfer aids	Bath lift/easy access bath tubs	Powered sit-to-stand devices	Total life hoist (ceiling)	# of safety equipment available	# Long-term Care facilites	Median injuries /100 employees
Safety equipment combination	√	√	√				√			4	10	8.33
	√	√	√	√	√	√		√		7	23	7.02
	√	√	√	√	√	√				6	9	6.97
	√	√	√	√		√	√	√		7	25	6.96
	√	√	√	√	√	√	√			7	13	6.96
	√	√	√	√	√			√		6	29	6.67
	√	√								2	10	6.64
	√	√	√				√	√		5	27	6.35
	√	√	√					√		4	66	6.30
	√	√	√	√		√		√		6	14	6.06
	√	√		√						3	16	5.98
	√	√	√	√	√					5	14	5.83
	√	√	√	√	√		√	√		7	43	5.71
	√	√	√	√			√	√		6	80	5.53
	√	√	√					√		4	19	5.43
	√	√	√	√	√	√	√	√		8	64	5.33
	√	√	√	√	√	√	√	√	√	9	11	5.01
	√	√	√							3	32	4.73
	√	√	√	√						4	56	4.67
	√	√	√	√		√	√			6	11	4.26
	√	√	√	√			√			5	42	3.63
	√	√		√			√			4	9	3.51
	√	√	√	√	√		√			6	19	3.23
	Other combinations										150	5.88
	Total (missing = 107)										**793**	**5.68**

PASSPORT clients – disenrollment? Predicting falls?

Figure O.1: Reasons for Disenrollment from the PASSPORT program



Source: Mehdizadeh, S., Nelson, I., Thieman, L. (2007). PASSPORT Consumer Eligibility. Scripps Gerontology Center, Miami University. PASSPORT consumers with an active service plan during October 1, 2004 to September 30, 2005. PASSPORT Information Management System (PIMS).

Motivation: b/c high proportion of people who disenroll from PASSPORT leave for nursing homes (much more costly care than in-home services), policymakers would naturally be curious about ways to intervene before such a move is necessary. Aside: PASSPORT = Medicaid funding for long-term care services at home or in the community.

In Ohio, avg. daily Medicaid NH reimbursement rate $164 /d;
Avg. PASSPORT home cost $48/d

By identifying those individuals who are most likely to leave for a nursing home, strategies could be employed to postpone use of this higher-cost alternative.

Goal: determining factors influencing nursing home admission

Study population: 4,654 individuals, of whom 325 (7.0%) disenrolled from PASSPORT to go to a nursing home [participants assessed from Oct. 03 – Sept. 04].

Predictor variables (26)

- Demographic variables (age, sex, race, marital status, length of time in PASSPORT)
- Presence/absence of pre-existing conditions (dementia, Parkinson's, Alzheimer's, stroke, depression, diabetes, emphysema, cancer, behavioral abnormalities, incontinence, cardio-pulmonary obstructive disease)
- Past-year's medical care (Admitted to a nursing facility, admitted to a hospital)
- Medication information (Yearly $ spent, # of total prescribed medications)
- Independence (Requires assistance administering medication, ability to care for oneself, presence of primary caregiver, number of ADL deficiencies[1], number of IADL[2] deficiencies)

---

[1] Activities of Daily Living (ADL) include: bathing, mobility, dressing, grooming, using the toilet, and eating.

[2] Instrumental Activities of Daily Living (IADL) include: shopping, meal preparation, performing heavy household chores, yard work, handling legal and financial matters, and laundry.

Methods

* Used a tree-structured classification algorithm called CRUISE (Classification Rule with Unbiased Interaction Selection and Estimation)

* final size of our tree is determined using an overfit-then-prune strategy - tree is grown very large until the size of each resulting subgroup is very small. To determine how much of the tree should be pruned, we estimate our classifier's out-of-sample performance using cross-validation, and the complexity of our final tree-structured classifier is selected using the 1-SE method.

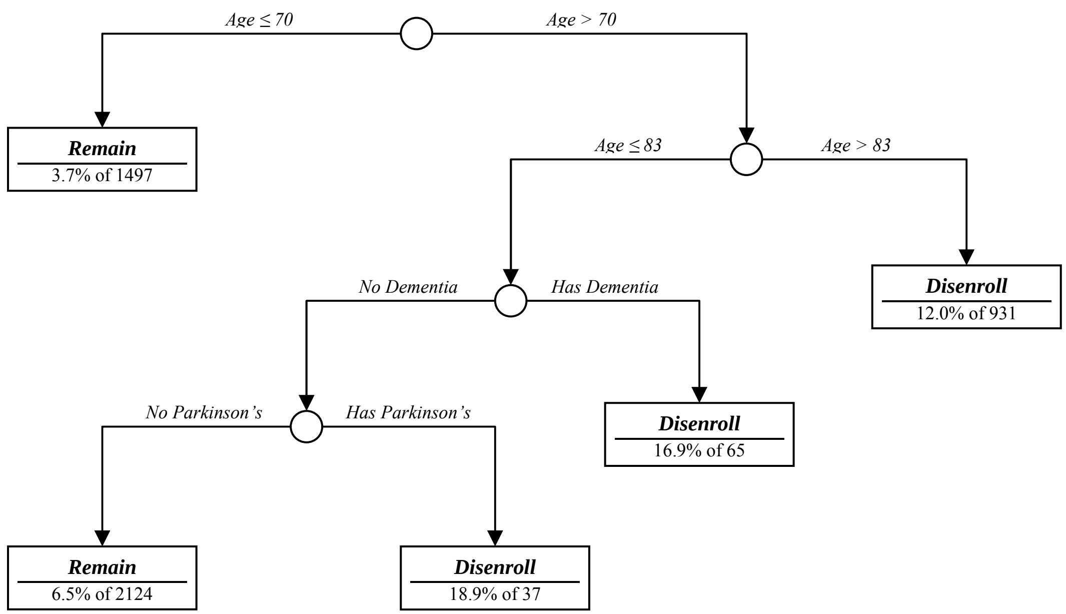
Details:

1. Index of node impurity minimized at each split (e.g. Gini index)

2. 10-fold cross-validation with simplest tree within 1 SE of the min(impurity) tree

   SE = sd of impurity measure over cross-validation samples for the min(impurity) tree

Figure O.2 – Classification tree for predicting PASSPORT disenrollment [correctly identifies 40.0% of those who disenroll and 79.1% of those who remain]



In summary, the following groups are identified as “high-disenrollment” groups:

- Age over 83
- Age 71-83 with either dementia or Parkinson’s disease

These groups are identified as “low-disenrollment” groups:

- Age 70 or younger
- Age 71-83 with neither dementia nor Parkinson’s disease

1. Evaluating resident care needs

* The health status and associated resource utilization for every nursing home resident is evaluated quarterly [and upon some change of status (e.g. readmission to the nursing home after a stay in a hospital)]

* Minimum Data Set (MDS) = standard assessment instrument

* MDS usually completed by a RN ("MDS or resident assessment coordinator") and includes input from other professionals and staff working at the nursing home (e.g. nutritionist, occupational therapist, etc.).

* MDS assessment yields a Resource Utilization Group (RUGS) category for each resident.

* Each level in this **44 level categorization** is then assigned a numeric score, the **case-mix score** ≈ amount of skilled nursing care needed by a resident.

* Case-mix scores are used to help set reimbursement levels for a facility so that facilities with higher case-mix scores receive higher levels of reimbursement.

* Motivation/concern: facilities might feel pressured to maximize their case-mix scores in order to generate the largest reimbursement.

- To evaluate, a sample of nursing homes in Ohio was selected, and independent assessors were sent to the sampled facilities
- Ratings of residents at each facility by the independent assessor were compared to ratings by staff at each nursing home
- **QUESTION:** are the two sets of raters comparable?
- **OPTIONS FOR ANALYSIS?:** Kappa agreement measures (44 categories and sparse counts even if simplify); Correlation of scores (again small sample sizes); Mean equality of scores; Bland-Altman displays …
- **Need to craft new method …**

2. Study and data structure

* we conducted a study in which independent assessors of an indicator of resource utilization, case-mix scores, were compared to nursing-home assessors.

* an independent team of Registered Nurse assessors were sent out to a sample of nursing homes in Ohio.

  - case-mix scores derived by this team of independent assessors for a sample of residents in each facility was compared to the case-mix scores generated by the nursing homes for the same residents.

Eight strata of nursing homes in Ohio formed using

- 4 geographic groups
- Above / below median care needs scores (in 1997)

Within each stratum, a simple random sample of 8 facilities was drawn (64 total facilities)

Each facility was assigned one of several independent assessors, who generated ratings for each of up to 10 residents previously rated by the facility.

229 residents from 39 facilities participated

- Facility participation rate ~ 60%

The independent assessor's rating (A) for each individual was compared to the facility's rating (F) for the same individual.

Each pair of ratings was classified as follows:

$F < A$ $F = A$ $F > A$

where "F>A" ↔ independent assessor (A) scores care needs of a resident as lower than the facility rater (F).

For a given independent assessor going to a particular nursing home.

Category	F < A	F = A	F > A
Count	$n_{F<A}$	$n_{F=A}$	$n_{F>A}$

where $1 \leq n_{F<A,ij} + n_{F=A,ij} + n_{F>A,ij} \leq 10$ for i=1,…,6 (raters) and j=1, …, n_i(residents)

Modeling the data

We model these data as being drawn from a trinomial distribution with class probabilities $\boldsymbol{\pi} = (\pi_{F<A}, \pi_{F=A}, \pi_{F>A})$:

$P(F < A) = \pi_{F<A}$ $P(F = A) = \pi_{F=A}$ $P(F > A) = \pi_{F>A}$

3. If gaming, then what is expected?

For a given independent assessor going to a particular nursing home.

$P(F < A) = \pi_{F<A}$ $P(F = A) = \pi_{F=A}$ $P(F > A) = \pi_{F>A}$

Assuming no systematic differences between a facility (F) and its independent assessor (A), we may formulate our null hypothesis as:

$$H_0: \pi_{F<A} = \pi_{F>A}$$

On the other hand, if there is a systematic tendency for the facility assessor (F) to generate a higher rating than the independent assessor (A), i.e. F>A more likely than F<A, we have:

$$H_A: \pi_{F>A} > \pi_{F<A}$$

4. So how can we test this?

4.1 Are the independent assessors the same?

4.2 What would you expect for each assessor?

4.3 Which facilities are flagged?

4.1 Are the independent assessors the same?

* permutation test (see, e.g., Good 2000) - facility labels of residents at the 39 facilities were randomly permuted.
* 5000 permutations of the data
* disagreement % was then calculated for each permuted data set
* observed disagreement percentage was compared to this permutation distribution and a permutation P-value was obtained

Details

1. $D_1 \ldots D_{39}$ are % disagreement observed in facilities with $n_1 \ldots n_{39}$ residents sampled, respectively.
2. $R_{i1} \ldots R_{in_i}$ ratings from each facility
3. Permute the labels "1", "2", … "39" associated with the $n_1+n_2+\ldots+n_{39}$ residents (constrained so that n_i get label "i") and let $\boldsymbol{D_i^*}$ = % disagreement associated with i^{th} facility in a permuted dataset. The $\{\boldsymbol{D_i^*}\}$ define a null distribution for comparing the observed % disagreement (assuming rater homogeneity).
4. p-value = $P(\boldsymbol{D_i^*} > D_i)$

4.2 What would you expect for each assessor?

Under H_0: same probability of independent assessor (A) generating a lower or higher rating than the facility assessor (F),

Category	F < A	F = A	F > A
Count	$\pi_{F<A}$	$\pi_{F=A}$	$\pi_{F>A}$

which for observed data

Category	F < A	F = A	F > A
Count	$n_{F<A}$	$n_{F=A}$	$n_{F>A}$

would yield

$$\hat{\pi}_0 = \hat{\pi}_{0,\boldsymbol{F}>\boldsymbol{A}} = \hat{\pi}_{0,\boldsymbol{F}<\boldsymbol{A}} = \frac{\boldsymbol{n}_{\boldsymbol{F}<\boldsymbol{A}} + \boldsymbol{n}_{\boldsymbol{F}>\boldsymbol{A}}}{2 \times \left(\boldsymbol{n}_{\boldsymbol{F}<\boldsymbol{A}} + \boldsymbol{n}_{\boldsymbol{F}=\boldsymbol{A}} + \boldsymbol{n}_{\boldsymbol{F}>\boldsymbol{A}}\right)}$$

4.3 Which facilities are flagged?

i. calculate $\hat{\pi}_0$ (or 1-2$\hat{\pi}_0$ for agreement probability) for each assessor (assuming heterogeneity in the independent assessors)

ii. calculate the $Pr(N_{F<A}=n_{F<A}, N_{F=A}=n_{F=A}, N_{F>A}=n_{F>A} \mid \hat{\pi}_0)$ for the observed data

iii. P-value = $\Sigma\, Pr(N_{F<A}=n_{F<A}, N_{F=A}=n_{F=A}, N_{F>A}=n_{F>A} \mid \hat{\pi}_0)$ where the sum is taken over all configurations of counts equal to or more extreme than observed in the direction of A<F.

For example, Assessor "Sp" had $\hat{\pi}$=0.314285714. Thus, we consider any set of resident ratings at a facility to be the realization of a multinomial random variable with $\underline{\hat{\pi}}$=(0.314, 0.371, 0.314)

Consider the P-value for a particular provider where n=10 residents where assessed where 6 cases with the assessor higher than the assessor (F<A), 1 tied case (A=F), and 3 cases where the assessor rated the resident higher than the facility (F<A).

F>A	F=A	F<A
6	1	3

More extreme would be cases shift towards "F<A", e.g.

F>A	F=A	F<A
6	2	2

In general, “more extreme” reflects a systematic tendency for the facility assessor (F) to generate a higher rating than the independent assessor (A), i.e. F>A more likely than F<A, we have:

$$H_A: \ \pi_{F>A} > \pi_{F<A}$$

The possible values as extreme, or even more so, than observed are

F>A	F=A	F<A	Probability
6	4	0	0.003851841
7	3	0	0.001862429
8	2	0	0.000590963
9	1	0	0.000111121
10	0	0	9.40257E-06
6	3	1	0.013037001
7	2	1	0.004727704
8	1	1	0.001000091
9	0	1	9.40257E-05
6	2	2	0.016546963
7	1	2	0.004000365
8	0	2	0.000423115
6	1	3	0.009334184
7	0	3	0.001128308

The probability for the first case (for example) is

$$\frac{10!}{6!4!0!} 0.314285714^6 0.371428571^4 0.314285714^0 = 0.003851841$$

The p-value (0.0567) is the sum of the above listed probabilities

Unfortunately, it isn't always this clear when defining "more extreme than observed."

Reformulating these hypotheses in terms of the ratio of disagreement probabilities, $\varphi = \pi_{F>A} / \pi_{F<A}$, we obtain:

$$H_0: \varphi = 1 \qquad \text{vs.} \qquad H_A: \varphi > 1$$

How do we test these hypotheses?

1. Estimate multinomial probabilities under H_0: $\boldsymbol{\pi_0}$
2. Calculate $P(n_{F<A}, n_{F=A}, n_{F>A} \mid \hat{\pi}_0)$ for the observed data
3. p-value = $\Sigma\, P(\mathbf{n} \mid \hat{\pi}_0)$, where the sum is taken over configurations of all counts at least as extreme as observed

Ordering the Trinomial Sample Space

Steps 1 and 2 are clear, but Step 3 is trickier. How do we define "as or more extreme" in the direction of F>A?

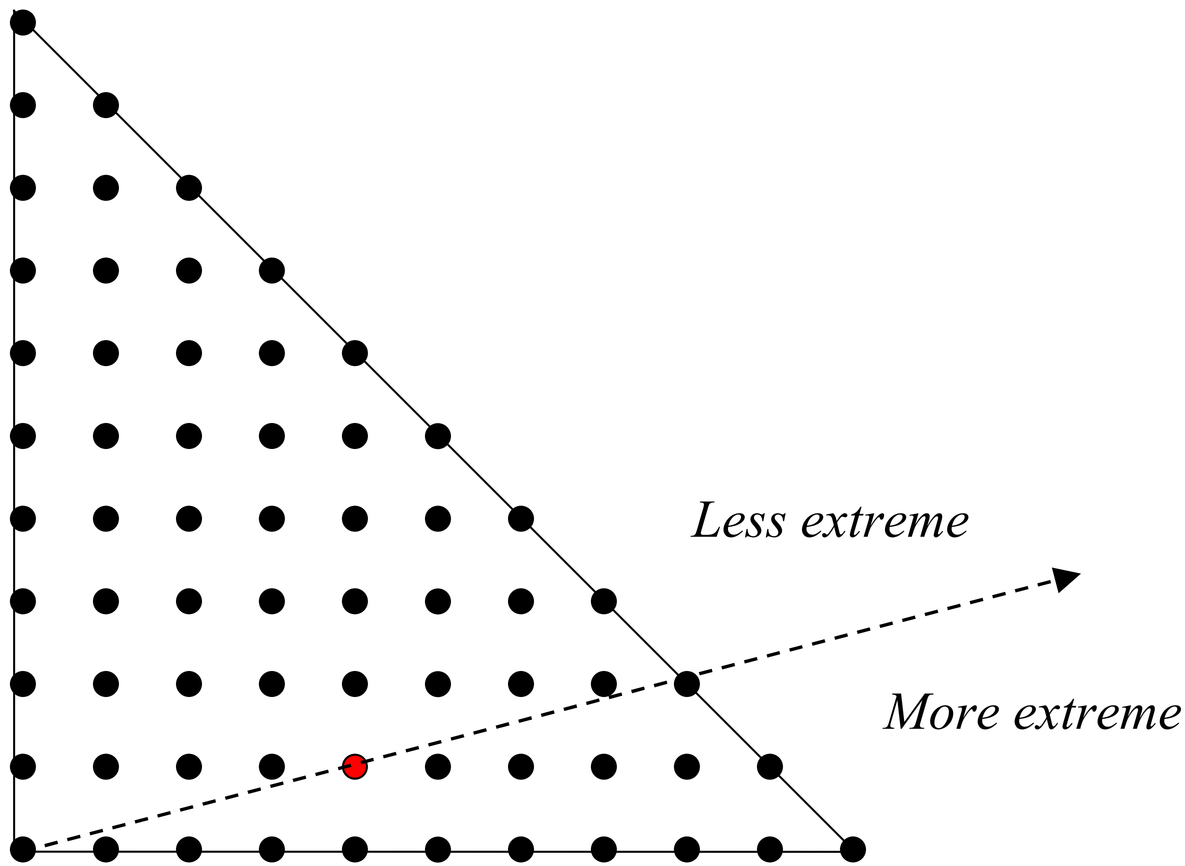
An ordering of the trinomial sample space (under H_0) is needed.

An intuitive possibility:

Calculate φ for each triplet in the sample space. Those that are greater than or equal to the value of φ calculated for the observed triplet are at least as extreme.

[note: $(n_{F>A}, n_{F=A}, n_{A>F})=(4,5,1)$ *may be written as is* $(n_{F<A}, n_{F=A}, n_{F>A})=(1,5,4)$ *or* $\varphi=4/1$]

Example: Observed triplet is $(n_{F<A}, n_{F=A}, n_{F>A}) = (1,5,4)$ *or* $\varphi = 4/1$



$n_{F<A}$

$n_{F>A}$

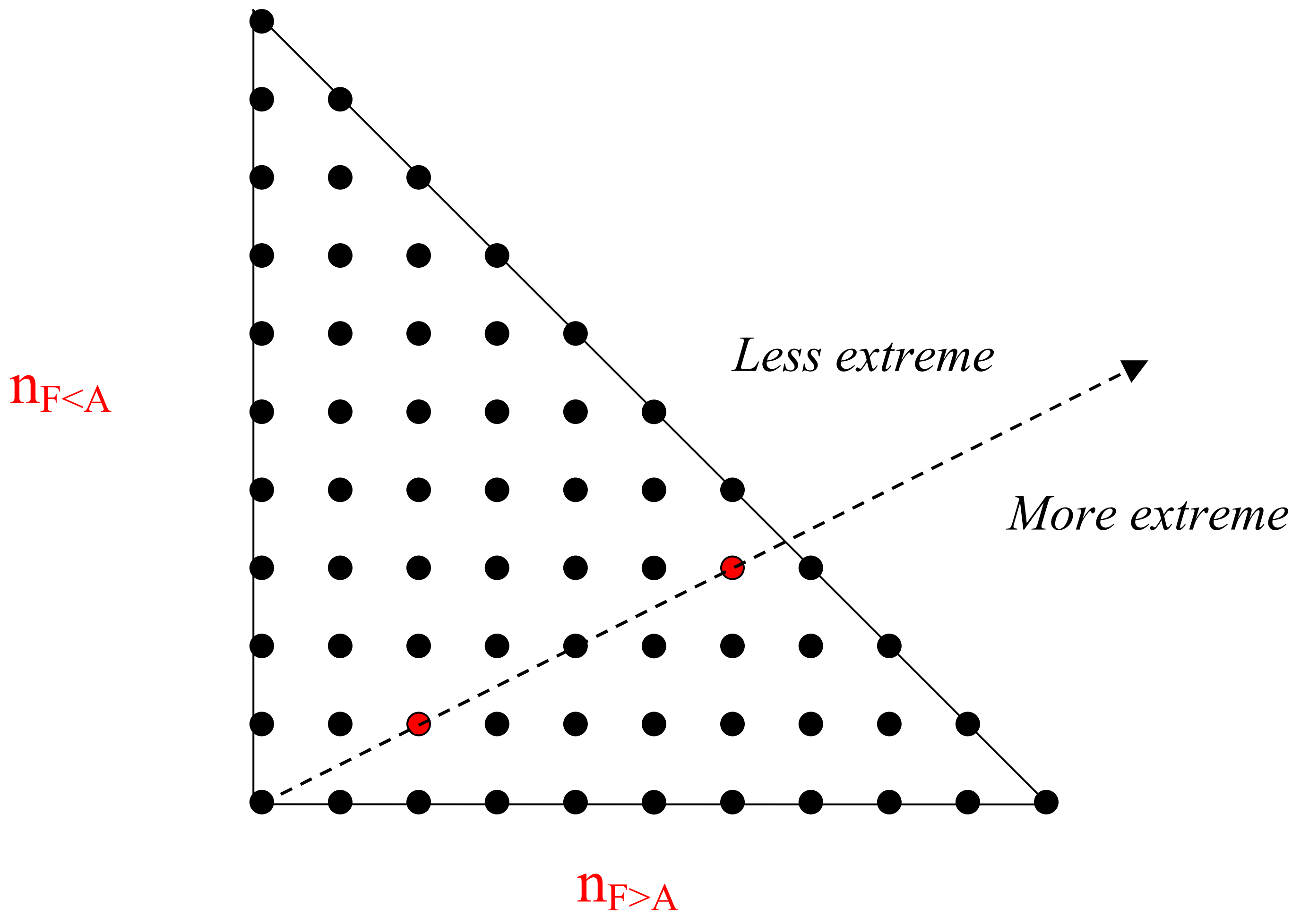
Comments:

1) *arrow/ray = evidence "equal" in terms of* $\varphi = \pi_{F>A} / \pi_{F<A}$ *to observed effect*

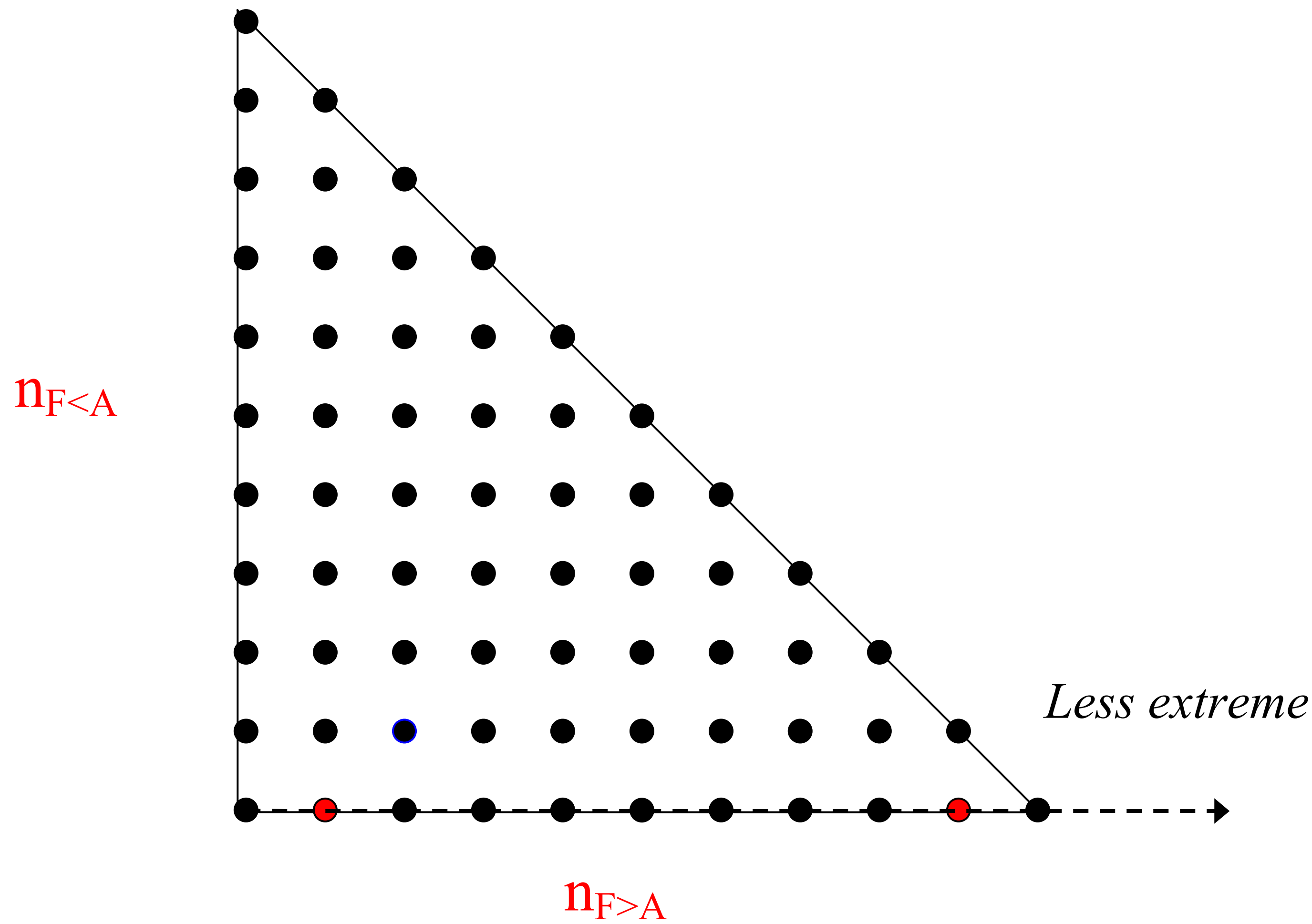
2) "More extreme" = $\{\mathbf{n} \text{ s.t. } \varphi > \varphi_{obs}\}$

3) "Less extreme" = $\{\mathbf{n} \text{ s.t. } \varphi < \varphi_{obs}\}$

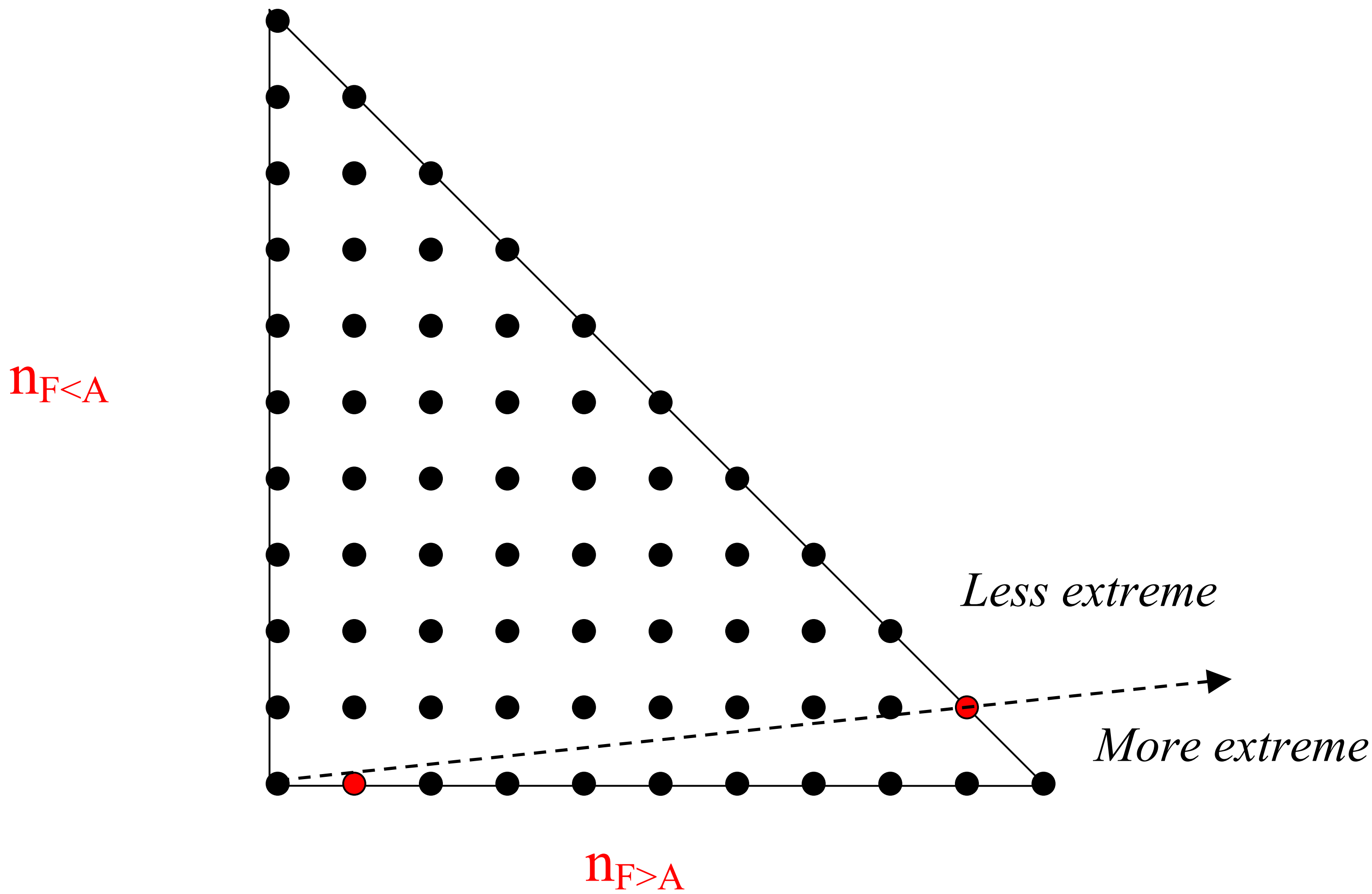
Problem 1: Is (1,7,2) [φ=2/1] really = extreme as (3,1,6) [φ=6/3]?



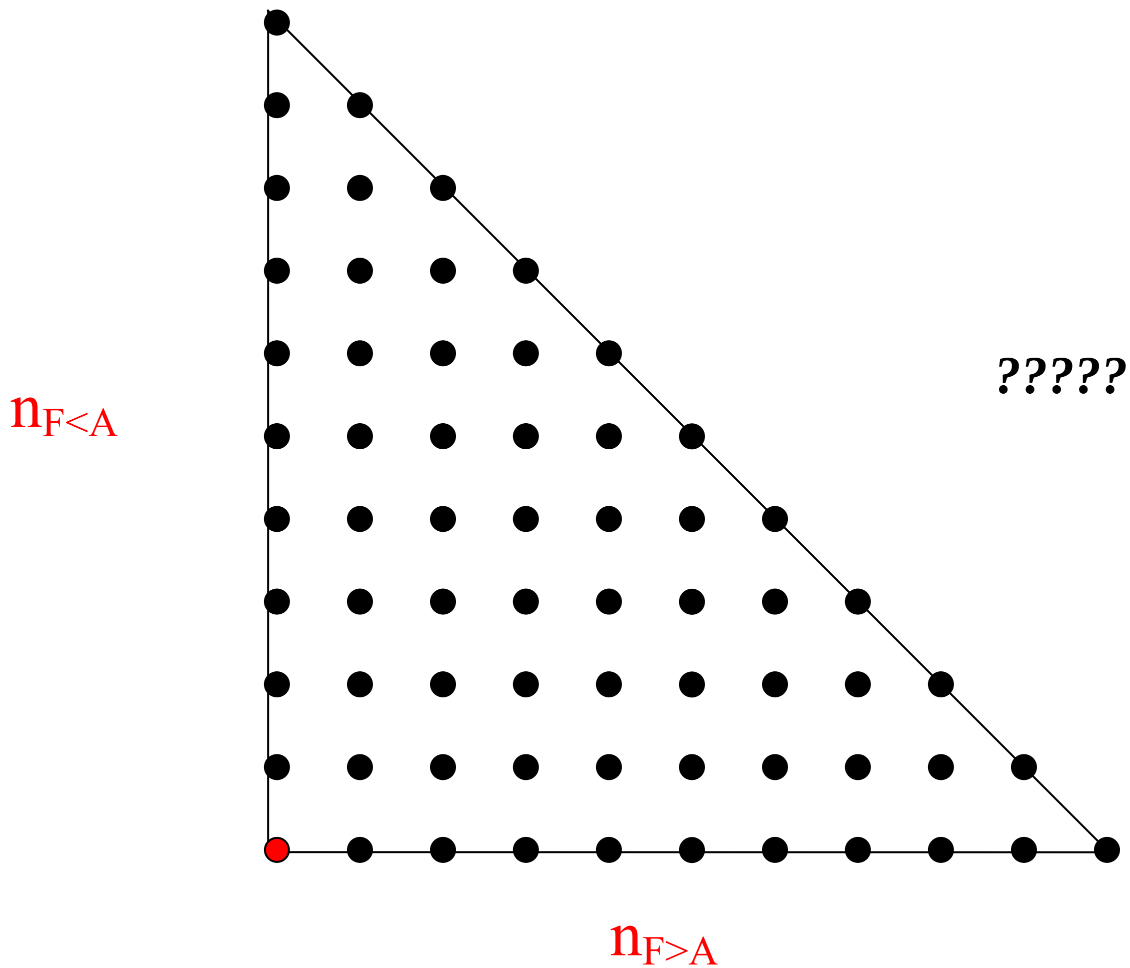
Problem 2: Is (0,9,1) [φ=1/0] really = extreme as (0,1,9) [φ=9/0]?



*Prob. 3:*Is (0,9,1) [φ=1/0] really *more extreme* vs. (1,0,9) [φ=9/1]?



Problem 4: What is more / less extreme than (0,10,0) [φ=0/0??]?



Some unintended consequences:

- The scenarios above do not seem reasonable.
- No hypothesis test will ***ever*** reject at the 5% level under the proposed ordering!

  All (0,y,z) configurations (with $z > 0$) are at least as extreme as any other sample point, therefore, the probabilities at each of these points is included in every p-value.

All other intuitive schemes (that we considered) retained at least one of these problems.

Bayesian Ordering

A priori, we have no knowledge of the trinomial probabilities $\pi_{F<A}$, $\pi_{F=A}$, and $\pi_{F>A}$, so we assume a uniform conjugate Dirichlet prior.

Given a trinomial outcome **n**, the resulting posterior density is

$$\boldsymbol{\pi} \mid \mathbf{n} \sim \text{Dirichlet}(1+n_{F<A}, 1+n_{F=A}, 1+n_{F>A}).$$

To order the sample space, we calculate the posterior probability of H_A ($\pi_{F>A} > \pi_{F<A}$) given each sample point – i.e. $P(\pi_{F>A} > \pi_{F<A} \mid \mathbf{n}) = P(\varphi > 1 \mid \mathbf{n})$.

- Higher probabilities are "more extreme"
- Lower probabilities are "less extreme"

Example: Order the two triples (3,0,1) [φ=3/1] and (2,2,0) [φ=2/0]. [Recall ($n_{F>A}$, $n_{F=A}$, $n_{F<A}$) and here index 1=F>A, 2=F=A, 3=F<A]:

$$P[\phi > 1 \mid (3,0,1)] = 120\left(\int_0^{0.5} \int_0^{\pi_3} \pi_1^3 \pi_3 \, d\pi_1 d\pi_3 + \int_{0.5}^{1} \int_0^{1-\pi_3} \pi_1^3 \pi_3 \, d\pi_1 d\pi_3 \right) = 0.8125 \text{, and}$$

$$P[\phi > 1 \mid (2,2,0)] = 180\left(\int_0^{0.5} \int_0^{\pi_3} \pi_1^2 (1-\pi_1-\pi_2)^2 \, d\pi_1 d\pi_3 + \int_{0.5}^{1} \int_0^{1-\pi_3} \pi_1^2 (1-\pi_1-\pi_2)^2 \, d\pi_1 d\pi_3 \right) = 0.875$$

- The probability of H_A is greater given (2,2,0) than given (3,0,1)

- (2,2,0) is a more extreme data point under H_0.

Example: A completely ordered sample space

Bayesian sample space ordering for a facility with 3 resident ratings:

Index	$n_{F>A}$	$n_{F=A}$	$n_{F<A}$	$P(\varphi > 1 \mid \mathbf{n})$
1	3	0	0	0.9375
2	2	1	0	0.875
3	1	2	0	0.75
4	2	0	1	0.6875
5	0	3	0	0.5
6	1	1	1	0.5
7	1	0	2	0.3125
8	0	2	1	0.25
9	0	1	2	0.125
10	0	0	3	0.0625

What do we conclude if a triplet of (2,1,0) is observed with a independent rater averaging 60% agreement with facilities?

Note that …

60% agreement implies $\pi_0 = (\pi_{F<A}=0.2,\ \pi_{F=A}=0.6,\ \pi_{F>A}=0.2)$

Calculating P-value

To calculate the p-value, we need only sum the probabilities of those observations "at least as extreme:"

Index	$n_{F>A}$	$n_{F=A}$	$n_{F<A}$	$P(\varphi > 1 \mid \mathbf{n})$	$P(\mathbf{n} \mid \pi_{F<A}=0.2$ $\pi_{F=A}=0.6, \pi_{F>A}=0.2)$
1	3	0	0	0.9375	**0.008**
2	2	1	0	0.875	**0.072**
3	1	2	0	0.75	0.216
4	2	0	1	0.6875	0.024
5	0	3	0	0.5	0.216
6	1	1	1	0.5	0.114
7	1	0	2	0.3125	0.024
8	0	2	1	0.25	0.216

9	0	1	2	0.125	0.072
10	0	0	3	0.0625	0.008

The p-value associated with a triplet of (2,1,0) is .080

5. What did we see?

5.1 Are the independent assessors the same?

Nope.

Assessors ranged from 22% to 90% agreement with the facility assessors.

A permutation test of Independent Assessor homogeneity yielded a P-value < 0.01 for 3 of the independent assessors.

5.2 What would you expect for each assessor?

Assessor	# Facilities	F>A	F=A	F<A	$\hat{\pi}$	% agree
K	5	5	14	1	0.15	70
Sm	5	2	14	2	0.11	78
Sp	5	11	14	11	0.31	38
T	7	19	29	3	0.22	56
W	13	2	71	6	0.05	90
Z	4	8	4	6	0.39	22

5.3 Which facilities are flagged?

Using methods described above, 5 of the 39 facilities surveyed (12.8%) exhibited ratings “too extreme” in favor of H_A at the 10% level of significance.

These results alone do not provide evidence of suspicious behavior; however, they provide a starting point for further investigation.

6. Summary

* Employed a multinomial prob. calculation as the basis of detecting disagreements between an independent assessor and a facility assessor

* differences could be attributed to differences between …

1. the facility and the independent assessor
2. independent assessors
3. facility assessors

[different data needed to tease this out …]

* can't separate an outlying independent assessor from a collection of nursing facilities that are systematically overestimating case-mix scores

* look at this as a screening tool to suggest further investigation of facilities

Reference:

Bailer, A.J., Noble, R.B., Straker, J.K., Noe, D.A. and Hughes M.R. (2008) Detecting systematic discrepancies in nursing home assessments of residents. *Health Services and Outcomes Research Methodology* 8: 19-30 and (published online: 28 November 2007).

Thank you for the invitation to come and speak with you!

Questions?