Rasch Analysis Assists BNHI with Revealing Most Misfitting Strings of Hospital Medical Service Information

Tsair-Wei Chien *Ou Lydia Liu* *Wen-Chung Wang*
Chi-Mei Medical Center *University of California, Berkeley* *National Chung Cheng University*
The corresponding author: smile@mail.chimei.org.tw

Abstract

The Bureau of National Health Insurance (BNHI) in Taiwan provides weekly hospital medical service information to the public. The authors downloaded datasheets from the BNHI website including growth rates of 569 hospitals comparing the counterpart-weeks of Year 2005 and 2004. Rasch analysis (1960)was conducted to estimate hospital growth rate abilities and to calibrate item difficulties using the computer program Winsteps. Results showed that the averaged growth rate of outpatient service among hospitals in Taiwan displayed a negative growth in 2005, either north sub-division bureau or the region hospital showed the most negative growth rate in its group respectively. The most unusual misfitting profile of any hospital generated by Winsteps was demonstrated to show up distorted growth on the specified item (week) and clearly expose the unexpected response in various week of outpatient service for putting more efforts on further detection.

1. Introduction

The Bureau of National Health Insurance (BNHI) in Taiwan weekly publishes hospital medical service information to the public since January of 2005. Rasch analysis of item response theory, which permits traits to be measured by comparisons within tests or within items, may assist BNHI with not only ranking levels of hospitals' growth rate but also revealing most misfitting strings of its service distortion under global budgeting system of health care.

How Rasch[1] analysis can easily and accurately assist BNHI with identifying the most misfitting strings of hospital medical service to the public and hospital managers for warning or tracing overgrowth will be demonstrated in this study.

2. Methods/Analysis

2.1. Subjects

The study sample was recruited from 569 hospitals' outpatient (outpatient department, OPD) growth rates resulted from the counterpart-weeks of year 2005 and 2004 on website datasheets of BNHI. Effective data were included in the study if they met the following criteria1: (1) no missing datum on each compared week; (2) at least 20 effective growth rates calculated in a year.

2.2. Procedure

Data were downloaded and collected in MS-Excel to form a 2x2 column-row datasheet representing 52 weeks and 477 hospitals respectively.

Using the concept of questionnaire survey with Likert [9] type multi-scoring scale, we divided each growth rate into 6 categories by the five thresholds of 3.5%, 7.0% and 0%, such as 1 representing less than –7%, 6 being more than 7% on each week's outpatient growth rate.

2.3. Measures

The 6-point items constitute the 52-week (indicator) inventory, which was developed to assess hospital outpatient growth activities.

2.4. Data Analysis

The analysis consisted of two parts. First, the unidimensional assumption of the evaluation inventory with appropriate categories was examined using Rasch analysis. The WINSTEPS computer program[2] was used to perform the Rasch analysis under the rating scale model.[3,4] Second, Rasch analysis with Winsteps computer program was applied to estimate hospital growth rate abilities, to calibrate item difficulties (i.e. weekly growth rate response), and to identify the most misfitting strings($p<.05$) for each hospital in order to investigate the most growing hospital and the most abnormal weeks in that hospital.

To examine unidimensionality of the evaluation inventory scale, the infit and outfit statistics were used to indicate whether the data fit the model's expectation. The infit mean square (MNSQ) is sensitive to unexpected behavior affecting responses to items near the hospital's ability measure; the outfit MNSQ is sensitive to unexpected behavior by persons on items far from the hospital's ability level.[4,5] MNSQ can be transformed to a t statistic, termed the standardized Z value (ZSTD), which follows approximately the t or standard normal distribution when the items fit the model's expectation.

In this study items with both infit and outfit MNSQ beyond the range of 0.5 and 1.5 were considered poor fitting. When items fit the model's expectation, the residuals (observed scores minus expected scores) should be randomly distributed. A factor analysis was conducted to verify whether any dominant component existed among the residuals. The unidimensionality assumption held if no dominant component was found.[6]

3. Results

A total of 477 hospitals participated in the this study. The characteristics of the study sample are shown in Table 1.

Table 1. Characteristics of the Hospitals (n=477)

Types	Med.Cntr	Region	District	Others	Total
Taipei	7(6.60%)	15(14.15%)	82(77.36%)	2(1.89%)	106
Northern	1(1.64%)	12(19.67%)	47(77.05%)	1(1.64%)	61
Central	3(3.16%)	11(11.58%)	80(84.21%)	1(1.05%)	95
Southern	3(3.95%)	13(17.11%)	54(71.05%)	6(7.89%)	76
Kao-Pin	3(2.44%)	12(9.76%)	106(86.18%)	2(1.63%)	123
Eastern	1(6.25%)	5(31.25%)	9(56.25%)	1(6.25%)	16
Total	18	68	378	13	477

3.1. Unidimensionality of Simplified Scale

When the Rasch model was applied to examine the 52 polytomous items, 9 misfit items (weeks) had outfit MNSQ statistics beyond the range of 0.5 and 1.5 , which were statistically significant at the 0.05 level. These 9 items were excluded from the rest of the analyses. A factor analysis on the residuals of Rasch-transformed scores showed no dominant factors. The first and second factors accounted for only 10.4% and 6.7% of the residual variance, respectively. These results indicated that there was a good model-data fit and that the assumption of unidimensionality held for these 43 polytomous items when assessed at this one year comparison of growth rates. We further examined the psychometric properties of the 43-item scale. The person measures (ranging from ┐-2.60 to 1.62) had a mean of ┐-0.35 and a SD of 0.63. The person reliability was 0.88 (which can be similarly interpreted as Cronbach’s), indicating that these items yielded reasonably good estimates for the hospitals.

The Figure 1 shows the map of persons and items along the continuum. The item difficulties were spread out across the hospitals, indicating that these items could well differentiate these patients.

The highest and lowest abilities of hospitals in growth rates shown in Figure 1 were the total of upper 7 and bottom 10 representing BNHI branches and hospital levels at first and second byte respectively.

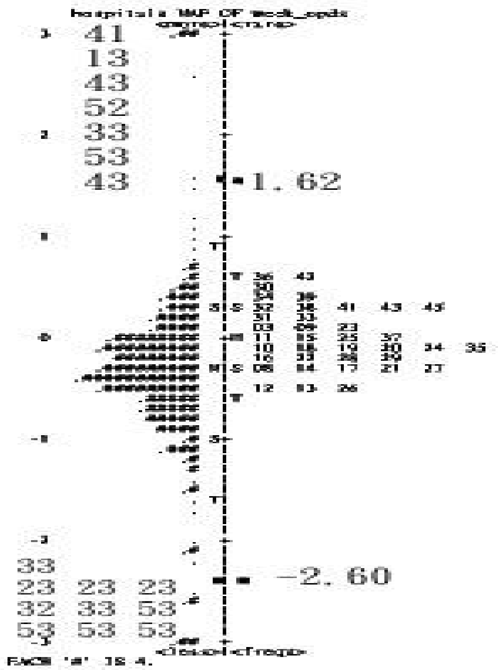

Figure 1. Item and person map under the Rasch model

3.2. Ranks of Hospitals’ OPD Growth Rates

The ranks of BNHI’s 6 region branches were highest 48.5(Eastern) to lowest 44.2(Northern) after measures in logit(M=0;SD=1) transformed into T-score (M=50; SD=10) to show up box plots from SPSS statistical software.

A box-plot of figure 3 shows that hospitals in Eastern region branch of BNHI are asymmetry. Some hospitals were shown as outliners or extremes with symbol o or * respectively in figure 2 and figure 3. Inter-quartile range of box-plot is appropriately set at 1.348 standard deviation occupied half of samples.[7] The probability of the outliners and extremes out from ± 1.5and ± 2 boxes in a box-pot is around .01 and .0005 respectively.[7] and stands for the growth rates much milder or severer needed a further investigation. Those hospitals out of boxes but inside outliners and extremes in box-plot are top or bottom 25% samples distributed in spread , symmetry or tail length referring to the shape of symbol 工.

Figure 4 shows the ranks of hospital levels are other-type (48.3), district (46.9), region (46.25) and medical center (45.45). There are two medical centers located beyond extremes and one out of outliners, which should be further inspected or investigated for reasons.

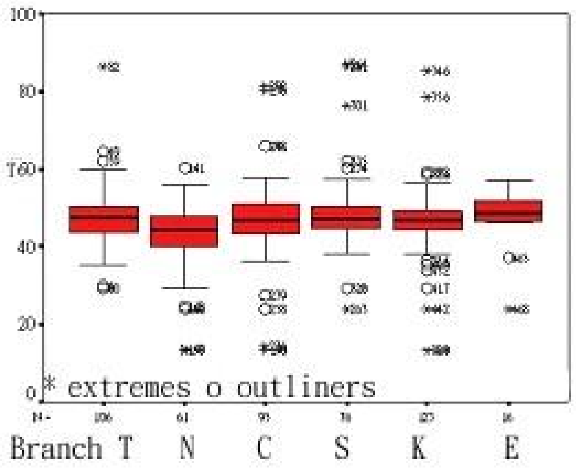

Figure 2. Growth rates between BNHI region branches.

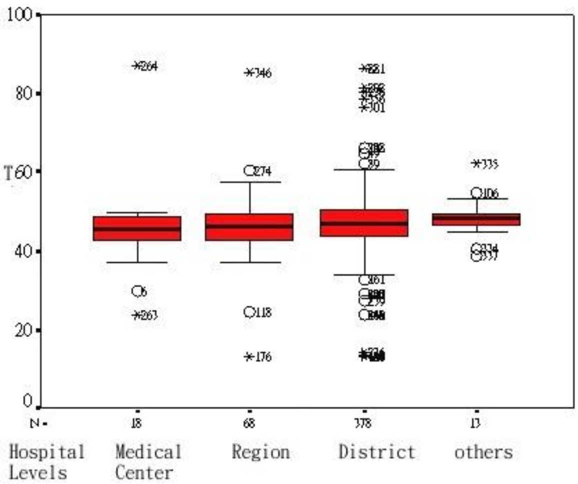

Figure 3. Growth rates between hospital levels.

3.3. Most Misfitting Strings of a Hospital

Besides knowing the abnormal BNHI branches or hospital levels in growth rates, most misfitging strings generated by Rasch analysis were provided with helping managers detect the abnormal weeks(items), which were statistically significant at the 0.05 level , in growth rates incurred in a hospital.

We showed a district hospital in Central region branch of BNHI in Figure 4. As for the demo hospital's ability at 1.01 and outfit MNSQ 4.3 bigger than the golden standard of 1.5[8], 3 weeks in Figure 4 shown up were most misfitting strings (p<.05) which indicated that the matched ability response at logit –1 reflecting the abnormal response in growth rates needed to be further inspected and understood.

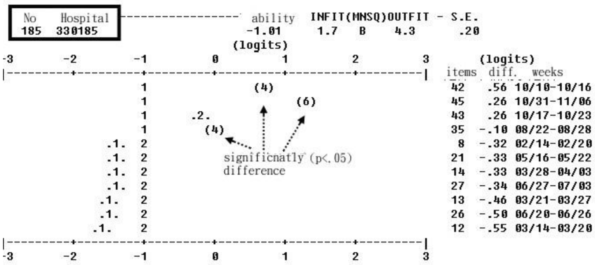

Figure 4. most misfitting strings shown in a hospital

3.4. Abnormal Phenomena on Internet

In order to allow all hospitals to easily monitor their abnormal items incurred to growth rates, we need to develop computer programs to be of practical usage.

An abnormal growth week on intra-hospital or inter-hospital was shown in Figure 5 and Figure 6 respectively.

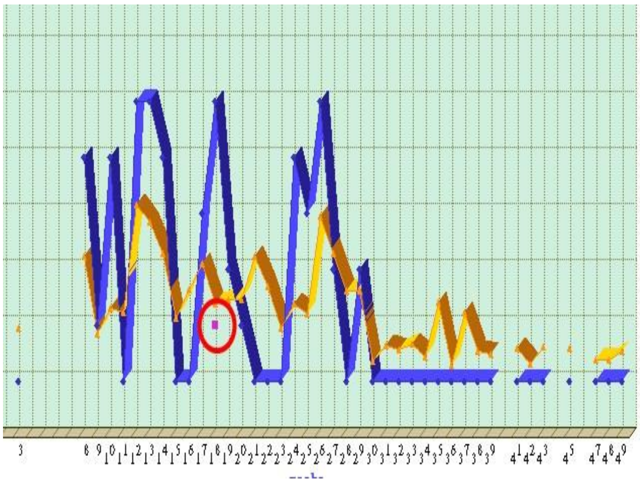

Figure 5. a circle shown abnormal on intra-hospital

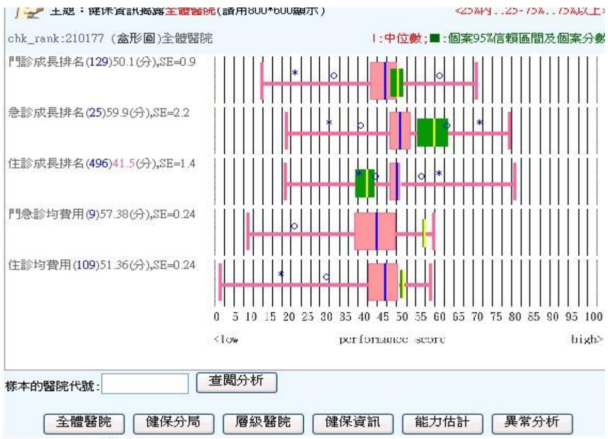

Figure 6. A box-plot shown abnormal on inter-hospitals

4. Discussion and Conclusion

Evaluation indices can provide a measurement tool with an enhanced range and sensitivity for assessing comprehensive hospital abilities in growth rates, which is especially useful when applied to hospitals with abnormal phenomena under global budgeting. Using Rasch analysis [1], this study found that such a measurement tool can be utilized to detect and delete noise growth weeks in order to make the measurement to be a unidimensional construct.

Because the items of the scale have been shown to measure comprehensive growth rates as a single, unidimensional construct for hospitals in a year, it is worth displaying the results on internet for mutual comparisons.

The Rasch-transformed scores can be viewed as interval-level measures [5,6], whereas the observed 6 categories raw scores are ordinal data. Since most statistical techniques assume that the data are at least on an interval scale, the Rasch-transformed scores of the growth evaluation inventory scale are suitable for comparing abilities between or within groups.[10] In addition, comparisons on internet should be widely adopted and developed through line-chart plots for easily monitoring the abnormal growth rates under global budgeting.

To use the new technique proposed in this article, the users can first obtain the knowledge of both psychometrics and computer science and then look up the ZSCORE [6] beyond ± 2 in a bit to detect the abnormal items incurred in a hospital.

There have been insufficient empirical investigations of the OPD growth rates between or within hospitals in such overall perspectives as this study offering an example not just of how research could be applied to inpatient or emergency departments for more detailed information of the phenomena of healthcare, but also of a regression model could be developed for the abilities of hospitals' OPD growth rates by factors of some predictive attributes in a future.

5. References

[1] G. Rasch, Probabilistic Models for Some Intelligent and Attainment Tests, Copenhagen, Denmark: Institute of Educational Research, 1960.

[2] J.M. Linacre, *WINSTEPS* [computer program]. Chicago, IL: http:// www.winsteps.com; 2006.

[3] D. Andrich.," A rating formulation for ordered response categories", Psychometrika, 43, 1978, pp. 561-573.

[4] B.D. Wright, G.N. Masters, Rating Scale Analysis, Chicago, Ill: MESA Press, 1982.

[5] B.D. Wright, M. Mok, "Rasch models overview", J Appl Meas, 1:1, 2000, pp.83–106.

[6] J.M. Linacre," Detecting multidimensionality: which residual data-type works best?", J Outcome Meas, 2:3, 1998, pp.266–283.

[7] Tukey, J. W., "Box-and-Whisker Plots." §2C in *Exploratory Data Analysis.* Reading, MA: Addison-Wesley, 1977, pp. 39-43.Retrieved at http://exploringdata.cqu.edu.au/box_norm.htm

[8] T.G. Bond, C.M. Fox, C.M.,Applying the rasch model: fundamental measurement in the human.sciences, Lawrence Erlbaum Assoc Inc,2001.

[9] R. Likert, "A Technique for the Measurement of Attitudes", Archives of Psychology, 140, 1932.

[10] S. E. Embretson, Item response theory models and spurious interaction effects in factorial ANOVA designs. Applied Psychological Measurement, 20:3, 1996, pp.201-212.