NHS England

Patent Reported Outcome Measures (PROMs)

An alternative aggregation methodology for case-mix adjustment

THE NHS CONSTITUTION
the NHS belongs to us all

NHS England INFORMATION READER BOX		
Directorate		
Medical	Operations	Patients and Information
Nursing	**Policy**	Commissioning Development
Finance	Human Resources	

Publications Gateway Reference:	**00213**
Document Purpose	Guidance
Document Name	Patient Reported Outcome Measures: An alternative aggregation methodology for case-mix adjustment
Author	NHS England Analytical Team
Publication Date	01 July 2013
Target Audience	Care Trust CEs, Foundation Trust CEs , Medical Directors, Directors of Nursing, NHS Trust CEs
Additional Circulation List	
Description	This paper sets out the new aggregation methodology for the national Patient Reported Outcome Measures (PROMs) programme
Cross Reference	Patient Reported Outcome Measures (PROMs) in England: The casemix adjustment methodology
Superseded Docs (if applicable)	Patient Reported Outcome Measures (PROMs) in England: The casemix adjustment methodology (aggregation section, para. 17 - 21)
Action Required	N/A
Timing / Deadlines (if applicable)	**N/A**
Contact Details for further information	PROMs Team Patients and Insight Analytical Team Area 6A Skipton House 80 London Road London SE1 6LH 0207 972 6144

Document Status

This is a controlled document. Whilst this document may be printed, the electronic version posted on the intranet is the controlled copy. Any printed copies of this document are not controlled. As a controlled document, this document should not be saved onto local or network drives but should always be accessed from the intranet

Patient Reported Outcome Measures (PROMs)

An alternative aggregation method for case-mix adjustment

First published: 1st July 2013

Prepared by: NHS England Analytical Team

Contents

Summary of Key Points

1. This paper puts forward an alternative aggregation methodology for the case-mix adjustment process for PROMs scores. This uses an additive approach that is often used in academic literature.

2. It has the advantage of weighting each patient individually providing more stable organisational level scores when patients' observed and expected scores are far apart.

3. It does not, however, give a bigger reward for treating patients with lower predicted scores (like the current method).

4. Comparisons with the current methodology used by NHS England show this alternative methodology produces consistent results.

5. Other aspect of the case-mix adjustment methodology remain unchanged

Description and Issues

1. This paper explores an alternative aggregation methodology (AAM) for case-mix adjusting PROMs scores. This suggestion addresses perceived weaknesses in the current methodology, namely that the resulting adjusted scores become unstable for certain patient scores[1]. The driver for this research has been continuous improvement, acting on feedback regarding rare situations where a patient's observed and expected scores can over influence organisational level scores.

2. For an alternative approach to aggregation to be acceptable, we believe it must give similar results to those under the current methodology while solving the problem outlined above.

3. The proposed methodology has arisen following discussions with academic colleagues who are familiar with using PROMs data and other data where standardisation across providers is necessary. It is also a suggestion made by the National Clinical Audit Advisory Group[2].

[1] When predicted EQ-5D index score is between zero and +/- 0.1 the ratio of actual score / predicted score becomes very large

[2] http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ab/documents/digitalasset/dh_123888.pdf

4. The current methodology uses the following equation to aggregate patient level scores to the provider level score. The full methodology can be found at the DH website[3]

$$Adj\ Q2_j = Q2_{National\ Avearge} \times \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\frac{\text{Actual PostOp Health}_{\mathrm{i}}}{\text{Predicted PostOp Health}_{\mathrm{i}}} \right)$$

Where:
AdjQ2$_j$ is the adjusted post-operative (Q2) score for provider j.
Actual PostOp Health$_i$ is the observed post-operative health for patient i
Predicted PostOp Health$_i$ is the post-operative health predicted by the model, given their case-mix

5. The alternative approach calculates the adjusted post-operative score for a provider as the national average plus the provider-level average difference between actual and predicted scores:

$$AdjQ2_j = Q2_{National\ Average} + \frac{1}{N} \sum_{i=1}^{N} \left(AQ2_{j,i} - PQ2_{j,i} \right)$$

Where:
AQ2$_{j,\,i}$ is the actual post-operative (Q2 score) for patient i treated at provider j
PQ2$_{j,\,i}$ is the predicted post-operative (Q2 score) for patient i treated at provider j

Advantages and drawbacks

6. This method has several advantages:
    a. It weights each patient equally, rather than on the basis of their predicted score.
    b. It avoids the need for rescaling the patient's scores to deal with negative EQ-5D scores.
    c. Individual patients' scores cannot influence the provider's score in such a way that makes it unstable.
    d. Like the current methodology, a provider whose patients report precisely the scores they were expected to achieve would have their score equal to the national average (i.e. the last two terms cancel.)

[3] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/152297/dh_133449.pdf.pdf

7. Although it is a lot more stable, this approach is not without its downsides. One drawback of this approach is that it does not place a greater emphasis on patients for whom the predicted score is lower. That is, it does not give the provider a greater reward for improving the health of a patient whose predicted score was very low. (See Appendix 1 for a worked example.)

8. Also, the current methodology generates a Relative Performance Factor (RPF) at both patient and provider level. This is a useful metric that shows the performance of the intervention in percentage terms and gives a clear indication of whether a patient's actual score is greater than their predicted score[4].

Implementation

9. Changing the aggregation method has a number of practical implications for calculating and comparing adjusted health gain scores. Firstly, to ensure the case-mix adjustment process doesn't change the national average health gain, adjusted scores are modified. This is currently done by dividing each patient's adjusted Q2 score by the national average RPF[5]. Under the new method, because the RPF is additive and not multiplicative, the average RPF[6] should be added to the adjusted Q2 scores, as opposed to dividing by it.

10. Secondly, we currently use funnel plots to compare providers performance. Providers' adjusted average health gain scores are compared to the mean national figure and control limits at two and three standard deviations from this. Because these statistics are based on national data, the new aggregation method will not have any meaningful effect on their values. (As illustrated below, the aggregation method only really affects figures at provider level where the provider is performing relatively few operations and one or more patients have extreme health gain scores).

11. Finally, it is currently necessary to cap some adjusted health figures due to the mathematical instability of the current aggregation method. This will not be necessary under the new method because of its superior stability (in fact, this is the main reason for introducing it).

---

[4] As the Relative Performance Factor is a ratio of the patient's Actual and Predicted scores, it is relatively straightforward to calculate this statistic, regardless of whether it is included in the case-mix adjustment model.

[5] This is a number very close to, but not equal to 1, e.g. in the case of the Oxford score for hip replacements in 2010/11 the figure was 1.006

[6] The RPF under additive aggregation has less meaning on its own than under multiplicative aggregation, but when combined with average national average q2 score it gives a measure of the magnitude and direction of a providers under/over performance.

Impact of AAM on provider level scores

12. To assess the differences between AAM and the current methodology, we analyse the composition of provider level scores from the PROMs programme for the years 2009-12.

13. Scores were adjusted for age, sex, IMD deprivation score, Q1 score, 12 patient reported comorbidities, ethnicity and procedure complexity[7]. Chart 1 compares provider level, adjusted Q2 scores using AAM (horizontal axis) and the current methodology (vertical axis) for the Oxford Hip Score[8]. Providers with less than 50 patients across 3 years of data are excluded from the graph. These providers have insufficient records for summary data to be presented. The scores are very similar and have a very high correlation coefficient of +0.98. The graph shows there is some difference in scores, the biggest difference in scores being no more than +/-0.8 points.

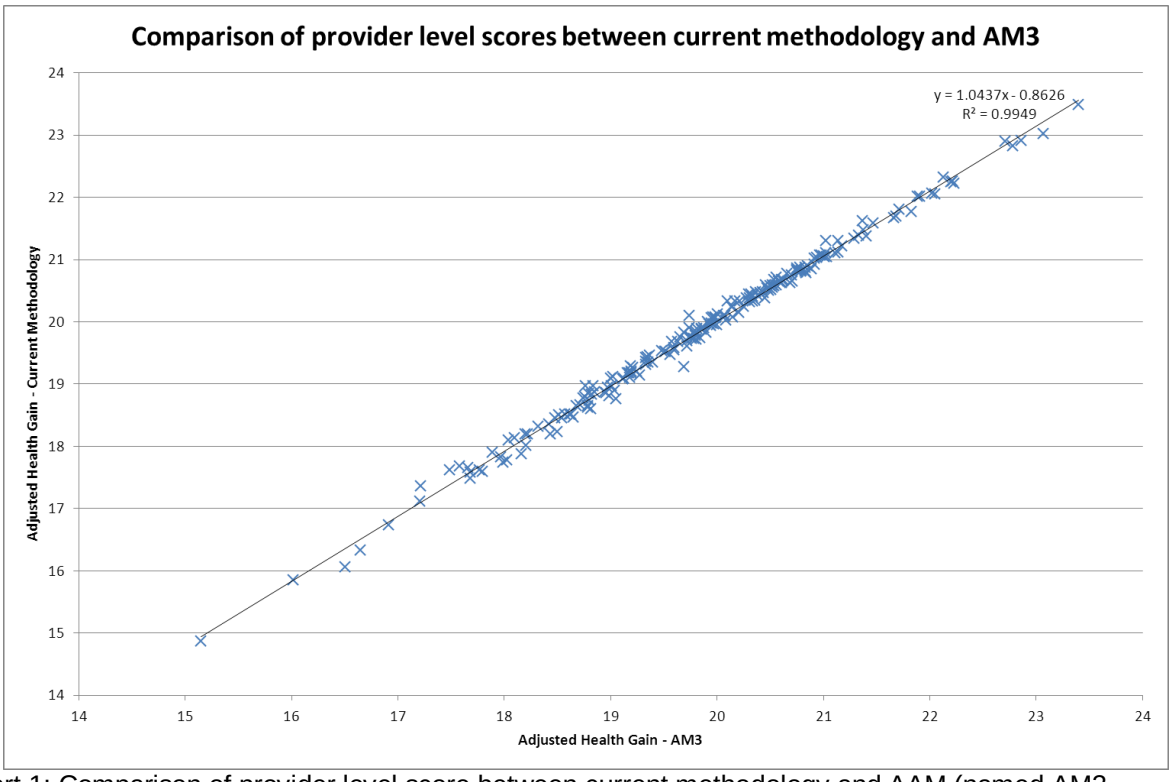

Chart 1: Comparison of provider level score between current methodology and AAM (named AM3 in chart) for Oxford Hip Score, 2009-12.

[7] A more thorough investigation of AM3 would include other factors in a risk adjustment model.
[8] The Oxford Hip Score is used as an illustrative example as this is more sensitive to patients' treatment.

Impact on identification of outliers

14. Changes to the methodology for calculating risk-adjusted scores will also have an impact on those providers who are considered to be outside of the control limits under the Department of Health's outlier policy[9]. Comparing the outlier status of providers under the two methodologies gives an indication of whether the two approaches are consistent. It also shows whether the new methodology would substantially change the status of any providers.

15. Using the adjusted scores from the previous section, the Table 1 shows the breakdown in outlier status between AAM (rows) and the current methodology (columns) for the Oxford Hip Score.

		Current Methodology					
	Outlier category	Positive 99%	Positive 95%	Control: Above	Control: Below	Negative 95%	Negative 99%
AAM	Positive 99%	19	0	0	0	0	0
	Positive 95%	0	16	1	0	0	0
	Control: Above	0	0	76	3	0	0
	Control: Below	0	0	3	48	0	0
	Negative 95%	0	0	0	2	15	2
	Negative 99%	0	0	0	0	1	23

Table 1: Outlier status for providers under current methodology (columns) and AAM (rows) for Oxford Hip Score, 2009-12.

16. Of the 209 providers with 50 or more procedures, 94% would remain in the same category under both methodologies. No provider changes by more than one category. The same tables have been produced for all the other instruments. These can be found in Appendix 2.

17. Table 2 shows the proportion of providers whose outlier categories are different under the current methodology and AAM. As a further investigation of stability, we analyse the datasets in-year as well, when there would be fewer patients per provider.

[9] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/151988/dh_133579.pdf.pdf

Procedure	Instrument	All	2009/10 only	2010/11 only	2011/12 only
Hip Replacement	Oxford Score	5.7%	4.2%	5.1%	4.2%
	EQ-5D Index	11.3%	7.9%	8.6%	6.0%
	EQ-5D VAS	7.4%	4.3%	5.3%	5.4%
Knee Replacement	Oxford Score	6.3%	9.2%	3.7%	4.1%
	EQ-5D Index	13.2%	14.0%	11.3%	12.2%
	EQ-5D VAS	7.0%	7.6%	8.9%	4.4%
Groin Hernia	EQ-5D Index	12.9%	6.6%	10.9%	12.1%
	EQ-5D VAS	12.0%	6.2%	13.4%	9.1%
Varicose Vein[10]	Aberdeen Score	4.7%	6.7%	10.7%	13.8%
	EQ-5D Index	17.5%	12.3%	14.0%	20.7%
	EQ-5D VAS	8.6%	7.8%	8.2%	6.9%

Table 2: Proportion of providers whose outlier statuses are different under current methodology and AAM

18. The instrument for which AAM has the greatest impact is EQ-5D index. This is expected and desirable as it is the instrument that had been identified as having the greatest instability issues. For all instruments and years, no more than 21% of providers would change outlier categories between the current methodology and AAM. The methodologies appear to give slightly different outlier categories to providers performing undergoing Varicose Vein procedures. However, DH and the HSCIC is aware of known problems in the Varicose Vein models, owing to less data being available to analyse and, for the Aberdeen Score, scores that operate in the opposite direction to all others.

19. These results, and the scatter plot suggest that, despite the mathematical and methodological differences between the current methodology and AAM, at provider level they do perform in a very similar way, with very few differences between the scores.

Case study

20. To investigate the impact of AAM in one extreme scenario, we look at a case study where the method of aggregation has caused adjusted post-operative scores to change substantially. In the 2011/12 PROMs data for a certain procedure, using the EQ-5D Index, provisional data suggests that Hospital A[11] has a health gain that is considerably outside all possible control limits.

[10] The reason why some of the varicose vein results are higher than others is that the number of providers of these procedures is much lower than for groin hernia or orthopaedics, typically around half the amount. Fewer providers therefore make percentage figures less stable.

[11] We have anonymised the case study in the interest of not unfairly singling out a specific provider.

21. The variation in the score is driven by one patient whose predicted score was close to zero but whose actual score was below zero. This creates a relative performance factor that distorts the provider's overall score.

22. Table 3 shows the adjusted post-op score for Hospital A under the current methodology and AAM. Comparing the difference in score between 'All patients' and 'without extreme patient' scores, there is more stability in the score under AAM compared to the current methodology. The difference in these scores can be explained by the mathematical differences discussed above.

			Adjusted Score	**Actual Score**
National Average 2011/12				0.704
Hospital A Score	**Current Methodology**	**All patients**	0.244	
		Without 'extreme' patient	0.597	
	Alternative Methodology 3	**All patients**	0.645	
		Without 'extreme' patient	0.647	
	2009/10 adjusted Q2 score		0.653	0.583
	2010/11 adjusted Q2 score		0.634	0.539
	2011/12 Q2 Score			0.568

Table 3: Adjusted and actual scores for Hospital A 2011/12 provisional data

23. Looking at previous years, the table above suggests that Hospital A have a caseload that is more complex than the national average, which is reflected in adjusted scores being higher than actual scores in 2009/10 and 2010/11. Assuming there are no substantial changes in the management and clinical resources to knee replacement, we would therefore expect a similar value for 2011/12.

Conclusion

24. This paper looks into an alternative methodology for case-mix adjustment of PROMs scores. The alternative aggregation approach uses an average additive approach to actual and predicted scores. This differs from the multiplicative approach used presently.

25. This approach has many benefits and one potential drawback, notably the fact it will not reward a provider more for treating patients with lower

predicted scores. It avoids the possibility for one patient to influence the score for a whole provider, as evidence from the case study highlights. Although no method is perfect the alternative method suggested has the benefit of being much more stable in circumstances where an extreme ratio is calculated due to the patient's observed and expected scores being far apart.

26. Comparing the performance of both metrics finds that providers' scores do not change significantly, and across all instruments the number of providers whose outlier status is different is not substantial.

27. On the basis of these findings, we recommend that the HSCIC adopts the Alternative Aggregation Methodology as an approach to case-mix adjustment through the national PROMs programme.

Dafydd Richards
NHS England, Patients and Insight Analytical Team

Appendix 1: Worked example

A1. This worked example supposes a provider treats four patients:

Patient	Actual Q2 score	Predicted Q2 score	Actual minus Predicted
1	10	50	-40
2	20	60	-40
3	30	70	-40
4	80	80	0
Q2 national average	90		
Provider's adjusted Q2 score	60		

Table A1: Example Q2 scores

A2. The predicted scores are an indication of the complexity of the patient's treatment. We can say, therefore, that Patient 1 has a more complex case-load than Patient 4. Ideally, a method of case-mix adjustment should be one that rewards a provider more for 'giving health' to Patient 1 compared to Patient 4.

$$AdjQ2_j = 90 + \frac{1}{4}(-40) + \frac{1}{4}(-40) + \frac{1}{4}(-40) + \frac{1}{4}(0)$$

$$AdjQ2_j = 90 - 10 - 10 - 0$$

$$AdjQ2_j = 60$$

A3. However, under this method, giving Patient 1 a health increase of 10 has the same impact on the provider's score as giving Patient 4 a health increase of 10. This impact differs to that under the current methodology, where the adjusted Q2 would be 48.6 with Patient 1's gain, and 47.0 with Patient 4's gain.

PATIENT 1

$$AdjQ2_j = 90 + \frac{1}{4}(-30) + \frac{1}{4}(-40) + \frac{1}{4}(-40) + \frac{1}{4}(0)$$

$$AdjQ2_j = 90 - 7.5 - 10 - 10 - 0$$

$$AdjQ2_j = 62.5$$

PATIENT 4

$$AdjQ2_j = 90 + \frac{1}{4}(-40) + \frac{1}{4}(-40) + \frac{1}{4}(-40) + \frac{1}{4}(10)$$

$$AdjQ2_j = 90 - 10 - 10 - 10 + 2.5$$

$$AdjQ2_j = 62.5$$

Appendix 2: Outlier categories for other PROMs instruments

HIP EQ-5D INDEX		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%	10						10
	Positive 95%		15	6				21
	Control: Above		2	72	2			76
	Control: Below			5	54	2		61
	Negative 95%				4	12	1	17
	Negative 99%					1	18	19
AM3	**Total**	10	17	83	60	15	19	204

HIP EQ-5D VAS		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%	7	1					8
	Positive 95%		16	2				18
	Control: Above		2	74	1			77
	Control: Below			3	66	2		71
	Negative 95%				4	16		20
	Negative 99%						9	9
AM3	**Total**	7	19	79	71	18	9	203

KNEE OXFORD		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%	19	1					20
	Positive 95%		18	2				20
	Control: Above		1	57	2			60
	Control: Below			1	66	1		68
	Negative 95%				1	11	3	15
	Negative 99%					1	22	23
AM3	**Total**	19	20	60	69	13	25	206

KNEE EQ-5D INDEX		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%	8	1					9
	Positive 95%	1	18	3				22
AM3	**Control: Above**		2	67	3			72

	Control: Below			7	65	2		74
	Negative 95%				2	9	3	14
	Negative 99%					3	10	13
	Total	9	21	77	70	14	13	204

KNEE EQ-5D VAS		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%	4	2					6
	Positive 95%		18					18
	Control: Above		2	74	2			78
	Control: Below			3	70	1		74
	Negative 95%				2	12		14
	Negative 99%					2	9	11
AM3	**Total**	4	22	77	74	15	9	201

GROIN EQ-5D INDEX		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%	1	1					2
	Positive 95%	1	6					7
	Control: Above		2	92	5			99
	Control: Below			7	78	2		87
	Negative 95%				7	5		12
	Negative 99%					2	1	3
AM3	**Total**	2	9	99	90	9	1	210

GROIN EQ-5D VAS		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%	1	1					2
	Positive 95%		7	2				9
	Control: Above		3	76	8			87
	Control: Below			7	92	1		100
	Negative 95%				1	8	1	10
	Negative 99%					1		1
AM3	**Total**	1	11	85	101	10	1	209

VEINS ABERDEEN	Current Methodology						
	Positive	**Positive**	**Control:**	**Control:**	**Negative**	**Negative**	**Total**

		99%	95%	Above	Below	95%	99%	
	Positive 99%	2						2
	Positive 95%		10	1				11
	Control: Above		1	43	1			45
	Control: Below				30	2		32
	Negative 95%					11		11
	Negative 99%						6	6
AM3	**Total**	2	11	44	31	13	6	107

VEINS EQ-5D INDEX		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%							0
	Positive 95%	1	4	2				7
	Control: Above			43	4			47
	Control: Below			7	33	3		43
	Negative 95%				1	2		3
	Negative 99%						3	3
AM3	**Total**	1	4	52	38	5	3	103

VEINS EQ-5D VAS		Current Methodology						
		Positive 99%	**Positive 95%**	**Control: Above**	**Control: Below**	**Negative 95%**	**Negative 99%**	**Total**
	Positive 99%							0
	Positive 95%		6	2				8
	Control: Above		1	41	1			43
	Control: Below			3	43	1		47
	Negative 95%					5		5
	Negative 99%					1	1	2
AM3	**Total**		7	46	44	7	1	105