

# Repeatability and reproducibility of upper eyelid measurements

K Boboridis, A Assi, A Indar, C Bunce, A G Tyers

## Abstract

**Aim**—The aim of this study was to assess the repeatability and reproducibility by physicians of upper lid measurements and to investigate the influence of clinical experience on the learning curve effect.

**Methods**—Both eyes of 22 outpatients were assessed for three basic measures of ptosis: marginal reflex distance (MRD) for upper and lower lids, upper lid skin crease (SC), and levator function (LF). Patients with variable eyelid positions were excluded. The patients were measured twice by a consultant and once by each of a clinical fellow, a specialist registrar, and a senior house officer in random order. Each observer was masked to their colleagues' results and followed a standard measurement protocol. Data were analysed using Bland-Altman plots.

**Results**—Consultant repeatability was high and consistent, the median difference between measures being 0 for each of the four parameters. Clinically acceptable reproducibility was shown in all measurements for even the least experienced physician and was particularly consistent for extreme observations. There was evidence of a learning curve effect.

**Conclusions**—These results suggest that interobserver and intraobserver variability in assessment of upper lid ptosis using a standard measurement protocol is low and clinically acceptable when the technique of assessment is standardised.

(Br J Ophthalmol 2001;85:99-101)

Adnexal Service,  
Moorfields Eye  
Hospital, London  
EC1V 2PD, UK  
K Boboridis  
A Assi  
A Indar  
A G Tyers

Glaxo Department of  
Epidemiology,  
Moorfields Eye  
Hospital, London  
EC1V 2PD, UK  
C Bunce

Correspondence to:  
Mr A G Tyers, Salisbury  
District Hospital, Salisbury  
SP2 8BJ, UK

Accepted 10 July 2000

Fundamental to the assessment of the patient and the choice of operation in ophthalmic plastic surgery is the measurement of certain parameters of the eyelids.<sup>1,2</sup> These include function in the levator muscle of the upper eyelid, the position of the eyelids with the eyes in the primary position, and the level of the upper lid skin crease.<sup>3-6</sup> Incorrect values may lead to an incorrect diagnosis or an inappropriate operation. It is our perception that

clinicians inexperienced in ophthalmic plastic surgery frequently make errors in these measurements. The purpose of this study was to test interobserver and intraobserver variability in such assessment. Clinicians of varying experience were included to allow examination of any learning curve effect.

## Patients and methods

Both eyes of 22 patients attending the ophthalmic plastic surgery clinic were assessed for the levator function (LF), the position of all four eyelids using the margin reflex distance method (MRD), and the level of the upper lid skin crease (SC).

A number of clinical conditions were included. Patients with variable eyelid position (for example, myasthenia gravis), inflamed eyelids, or photophobia were excluded. For each measurement the clinician sat in front of the patient at the same level and both looked in the primary position.

To measure LF the patient's eyebrow was stabilised by pressure exerted with the examiner's thumb. The patient was requested to look fully up, then fully down, while the excursion of the eyelid was measured against a ruler. This was repeated three times and the average measurement was recorded.

To measure MRD the patient was requested to look at a light source (pen torch) and the distances from the corneal light reflex to the upper eyelid and to the lower eyelid were recorded.

To measure the level of the SC the patient was asked to look down, the upper lid skin fold was gently raised if necessary, and the distance of the skin crease from the eyelid margin was recorded. The values were estimated to the nearest full millimetre.

Four clinicians took part in the study: a consultant, a fellow in ophthalmic plastic surgery, a specialist registrar, and a senior house officer. Both junior clinicians were attached to the Adnexal Service.

Each clinician assessed each patient in random order, the consultant measuring patients twice. Observers were masked to their fellows' assessments. The collection of data

Table 1 Range of eyelid measurements against clinician

	Consultant		Clinical fellow		Specialist registrar		Senior house officer	
	Right	Left	Right	Left	Right	Left	Right	Left
MRD (upper)	0, 5	0, 4	1, 5	1, 4	1, 6	-1, 4	1, 6	1, 5
MRD (lower)	4.5, 9	4.5, 8	4, 8	4, 7	5, 9	4, 7	4, 8	4, 8
SC	4.5, 10	4, 10.5	4, 11	4, 10	5, 10	4, 10	3, 10	3, 10
LF	5, 17	8, 17	5, 16	8, 17	5, 17	10, 17	6, 20	7, 20

Range (min, max) of measurements in mm.

MRD = marginal reflex distance; SC = skin crease; LF = levator function.

Table 2 Consultant repeatability in eyelid assessment

	Right	Left
MRD (upper)	0 (1)	0 (1)
MRD (lower)	0 (1)	0 (1)
SC	0 (1)	0 (1)
LF	0 (2)	0 (2)

Values are median (maximum difference) in mm.

MRD = marginal reflex distance; SC = skin crease; LF = levator function.

Table 3 Clinician reproducibility in eyelid assessment

	Clinical fellow		Specialist registrar		Senior house officer	
	Right	Left	Right	Left	Right	Left
MRD (upper)	0 (2)	0 (1)	0 (2)	0 (1.5)	0.5 (2)	0.25 (3)
MRD (lower)	0 (1)	0 (1.5)	0 (3)	0 (1.5)	0 (1)	0.25 (2)
SC	0.5 (3)	0.5 (3)	-0.5 (3.5)	-0.5 (3.5)	-1 (3)	-0.75 (4.5)
LF	0 (4)	1 (3.5)	1 (4.5)	2 (5)	1.5 (5)	2 (5)

Values are median (maximum difference) in mm.

MRD = marginal reflex distance; SC = skin crease; LF = levator function.

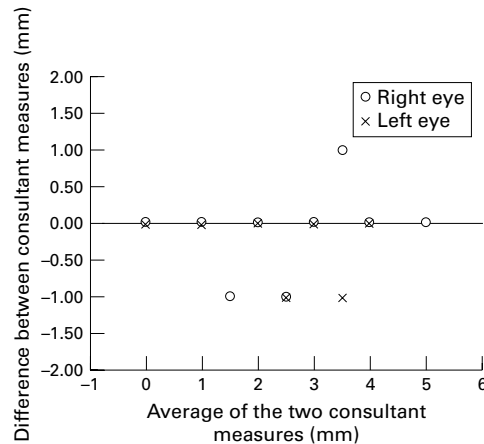


Figure 1 Bland-Altman plot of consultant repeatability for measurement of marginal reflex distance for upper eyelid.

was spread over a 3 month period so that it would be possible to assess any trend in the accuracy of the readings obtained by any of the observers.

#### ANALYSIS OF DATA

Table 1 presents the range of measures on each parameter made by each clinician. The data were analysed with the graphical method described by Bland and Altman for method comparison studies.<sup>7</sup> To test the intraobserver accuracy the difference between the two sets of measurements recorded by the consultant was plotted against the average of the two for each parameter. The average of the two sets of measurements made by the consultant was regarded as the best estimate of the true value. To test the interobserver accuracy the average value and the difference between this value and the results of the other observers were plotted as above. To assess whether there was any learning curve the position of the patient in the series and the time from the beginning of the study were noted.

Tables 2 and 3 provide information on repeatability and reproducibility, respectively. The right and left eyes were analysed separately.

#### Results

Table 1 shows that the measurements lay within the expected ranges for the parameters measured.

#### REPEATABILITY

Table 2 illustrates that consultant repeatability was high for all four parameters. On average, the second measure was the same as the first

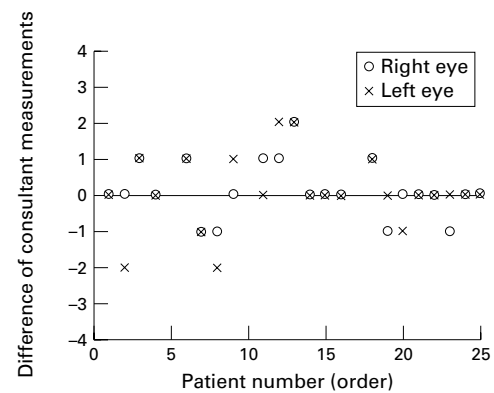


Figure 2 Repeatability of consultant measures of levator function against time.

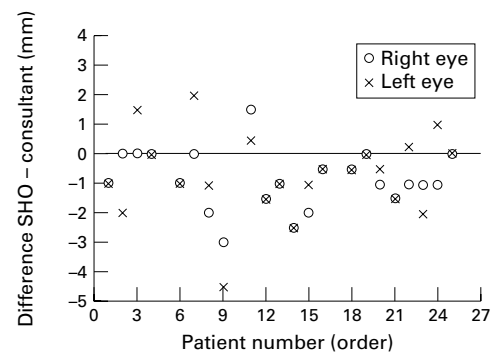


Figure 3 Reproducibility of skin crease measurement between senior house officer (SHO) and consultant with time.

(shown by the median difference being 0 in all case) and never differed by more than 2 mm. There was little evidence that repeatability varied with the size of the measurement (Fig 1) or with time (Fig 2).

#### REPRODUCIBILITY

Table 3 illustrates high reproducibility between the clinical fellow and the consultant. On average, there was no difference between their measures of MRD and a difference of just 0.5 mm for SC and 1 mm for LF. There was only slightly less reproducibility between the specialist registrar and the consultant with, on average, no difference between their MRD measures and a difference of 0.5 mm for SC. Table 3 does show, however, that the specialist registrar tended to record LF at 2 mm greater than the consultant and recorded a value 5 mm greater for one patient. There was slightly poorer reproducibility between the consultant and the senior house officer with, on average, differences of 0.5 mm, 1 mm, and 1.5 mm for MRD (upper), SC, and LF (right eye), respectively. While greater absolute differences between measures were seen in LF assessment, the range of acceptable values was greater for LF than for SC and MRD, so the differences were proportionately of similar clinical significance.

Figure 3 illustrates the slight increase in SC reproducibility over time. While in general there seemed little indication from these data of any variability in reproducibility over the parameter ranges, there is some suggestion that

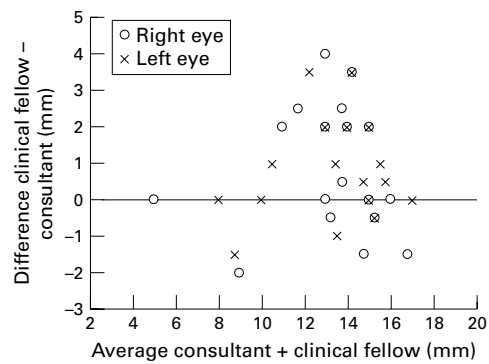


Figure 4 Reproducibility of levator function measurement between clinical fellow and consultant.

agreement between clinicians is greater at the extremes of LF (Fig 4).

### Discussion

This study suggests that interobserver and intraobserver variability in upper eyelid ptosis assessment, when conducted in a standardised

fashion, is modest and clinically acceptable, particularly in clinicians of greater experience. We have found some evidence of learning curve effects, both short and long term, and there is some suggestion of greater agreement at extremes of LF. This seems intuitive as a clinician may well check unusual observations more thoroughly than those that fall within typically encountered ranges.

- 1 Beard C. Examination of the ptosis patient. In: *Ptosis*. 3rd ed. St Louis: Mosby, 1981:76–83.
- 2 Collin JRO. Ptosis. In: *A manual of systematic eyelid surgery*. 2nd ed. London: Churchill Livingstone, 1989:41–72.
- 3 Beard C. Blepharoptosis: clinical evaluation and therapeutic rationale. *Int Ophthalmol Clin* 1970;10:97–115.
- 4 Fox S. Preoperative examination. In: *Surgery of ptosis*. 1st ed. Baltimore: Williams & Wilkins, 1980:25–30.
- 5 Karesh J. Diagnosis and management of acquired blepharoptosis and dermatochalasis (review, 65 refs). *Facial Plastic Surg* 1994;10:185–201.
- 6 Small R, Sebaton N, Burrows D. The measurement and definition of ptosis. *Ophthalmic Plastic Reconstr Surg* 1989;5:171–5.
- 7 Bland J, Altman G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;307–10.



## Repeatability and reproducibility of upper eyelid measurements

K Boboridis, A Assi, A Indar, et al.

*Br J Ophthalmol* 2001 85: 99-101

doi: 10.1136/bjo.85.1.99

---

Updated information and services can be found at:

<http://bjo.bmj.com/content/85/1/99.full.html>

---

### References

*These include:*

This article cites 2 articles

<http://bjo.bmj.com/content/85/1/99.full.html#ref-list-1>

Article cited in:

<http://bjo.bmj.com/content/85/1/99.full.html#related-urls>

### Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

### Topic Collections

Articles on similar topics can be found in the following collections

[Public health](#) (376 articles)

---

### Notes

---

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>