Classifying Stability of Misalignment in Children With Esotropia Using Simulations

B. Michele Melia, ScM; Jonathan M. Holmes, BM, BCh; Danielle L. Chandler, MSPH; Stephen P. Christiansen, MD; for the Pediatric Eye Disease Investigator Group

Objective: To determine the sensitivity and specificity of several classification rules for stability and instability of angle in childhood esotropia.

Methods: We conducted 10 000 Monte Carlo simulations of participants with no actual change in angle of esotropia during follow-up, where "observed" changes in ocular alignment were sampled from a distribution of measurement errors for the prism and alternate cover test. Additional simulations were conducted for a range of "true" changes (1.0, 2.5, 4.2, 5.0, 7.5, and 10.0 prism diopters [PD] per visit) with up to 10 follow-up visits. We then estimated sensitivities and specificities for specific rules for retrospectively classifying stability (all measurements within 0, 5, 10, or 15 PD) and instability (≥2 measurements differing by ≥10 PD, etc) across a fixed number of visits. Results were extended to classifying ocular alignment stability and instability prospectively based on a varying number of measurements.

Results: For a series of 4 measurements, the rules that optimized sensitivity and specificity were "all measurements within 5 PD" for stability and "at least 2 measurements differing by 15 PD or more" for instability. For a series of 3 measurements, all 3 measurements needed to be identical to confirm stability.

Conclusions: We derived definitions of stability and instability in childhood esotropia using estimates of actual measurement error that may be useful for clinical practice and for future clinical studies of esotropia.

Arch Ophthalmol. 2010;128(12):1555-1560

Author Affiliations: Jaeb Center for Health Research, Tampa, Florida (Mss Melia and Chandler); Department of Ophthalmology, Mayo Clinic College of Medicine, Rochester, Minnesota (Dr Holmes); and Department of Ophthalmology, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts (Dr Christiansen).
Group Information: A link to a list of the Pediatric Eye Disease Investigator Group members appears on page 1560.

ONE FACTOR THAT MAY affect the timing of strabismus surgery is whether the magnitude of the angle of misalignment seems to be changing (instability) or not changing (stability) with time. Although stability has been defined by some authors for data analysis,[1-4] previous definitions do not account for the presence of measurement error in strabismus assessment. Measurement error can have a large effect on the accuracy of classification when the classification is based on a cutoff point for a continuous measurement.[5] The choice of the cutoff point itself also affects misclassification. Real data are not directly useful for determining the extent of misclassification because real data include measurement error, and, hence, the true value and the true classification are unknown. As a result, simulated data are commonly used to determine the impact of measurement error on misclassification.

In the present study, we used Monte Carlo simulations to quantify the effect of measurement error. We explore the effect of several rules using differing cutoff points for defining stability and instability on misclassification by calculating sensitivity and specificity. In each simulation, ocular alignment data are generated using a statistical model in which the "true" angle of misalignment and changes in the true angle are specified as part of the simulation and, thus, are known. The simulated ocular alignment data also include "observed" data that incorporate estimates of measurement error into the true data. The estimates of measurement error (4.2 prism diopters [PD] for angles >20 PD and 1.7 PD for angles ≤20 PD) were based on results from a previous test-retest study[6] and did not vary by age. Based on the present findings, we make recommendations regarding rules for prospective classification of stability and instability in children with esotropia.

METHODS

MONTE CARLO SIMULATIONS

Data for a true angle and an observed angle at 10 time points were simulated for 10 000 hypothetical participants with moderate esotropia (baseline angle within 20-50 PD) and no change in angle over time according to the procedures

©2010 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 05/23/2013

Table 1. Pragmatic Rules Evaluated for Angle Stability or Instability

Rules for stability based on fixed No. of measurements classified retrospectively
Rule 1: If all measurements are within 0 PD, classify as stable; otherwise, classify as not stable
Rule 2: If all measurements are within 5 PD, classify as stable; otherwise, classify as not stable
Rule 3: If all measurements are within 10 PD, classify as stable; otherwise, classify as not stable
Rule 4: If all measurements are within 15 PD, classify as stable; otherwise, classify as not stable
Rules for instability based on fixed No. of measurements classified retrospectively
Rule 1: If ≥2 measurements differ by ≥25 PD, classify as unstable; otherwise, classify as not unstable
Rule 2: If ≥2 measurements differ by ≥20 PD, classify as unstable; otherwise, classify as not unstable
Rule 3: If ≥2 measurements differ by ≥15 PD, classify as unstable; otherwise, classify as not unstable
Rule 4: If ≥2 measurements differ by ≥10 PD, classify as unstable; otherwise, classify as not unstable
Rules for stability based on prospective measurement
Rule 1: If the last 3 consecutive measurements are all within 0 PD, classify as stable; otherwise, continue follow-up
Rule 2: If the last 4 consecutive measurements are within 5 PD, classify as stable; otherwise, continue follow-up
Rule 3: If the last 3 consecutive measurements are within 0 PD or the last 4 measurements are within 5 PD, classify as stable; otherwise, continue follow-up

Abbreviations: PD, prism diopters.

in the Appendix (available at http://www.archophthalmol.com). This process was repeated to simulate data for participants with a constant, predefined amount of change in angle (1.0, 2.5, 4.2, 5.0, 7.5, and 10.0 PD) between successive time points. The final data set contained 130 000 simulated participants, each of whom had a constant, predefined amount of change in angle between successive visits that ranged from −10.0 to 10.0 PD, including 0 PD (no change over time).

CLASSIFICATION RULES FOR STABILITY AND INSTABILITY

We considered 2 situations in which classification of stability and instability of ocular alignment is of interest: (1) when each participant has a fixed number of measurements and stability is being classified retrospectively, such as for a research study, and (2) when the participant is being observed prospectively for stability, such as in clinical practice. In the latter case, the number of measurements is not fixed in advance; at any follow-up time a decision may be made to classify the participant as stable based on the available measurements or to continue follow-up. The 2 situations require different classification rules.

RULES FOR RETROSPECTIVE CLASSIFICATION BASED ON A FIXED NUMBER OF MEASUREMENTS

Clinicians commonly think of changes in strabismus angle according to whether the measurement exceeds specific steps in the prism sets used to measure the misalignment. For moderate angles of strabismus (20-50 PD), commonly available prism sets have 5-PD increments. Hence, we considered classification rules that used multiples of 5 PD as the cutoff points for classification. The rules that we evaluated are listed in **Table 1**.

Although in truth a participant's alignment is either stable or unstable, the stability rules classify a participant's alignment as either stable or not stable and the instability rules classify a participant's alignment as either unstable or not unstable. Note that the definitions of stable and unstable are mutually exclusive; however, a classification of not stable (using the stability rule) is not the same as a classification of unstable (using the instability rule). Likewise, a classification of stable is not the same as a classification of not unstable.

RULES FOR PROSPECTIVE CLASSIFICATION

As with the rules for classifying stability and instability retrospectively based on a fixed number of measurements, rules for classifying stability and instability prospectively were based on using multiples of 5 PD for cutoff points as listed in Table 1.

STATISTICAL ANALYSIS

For analysis, the true alignment of participants with no change (0 PD) in true angle over time was considered stable, and the true alignment of those with any amount of change in true angle was considered unstable. The interval need not be specified; the rules can be applied to any evenly spaced measurements, with the results applying over the interval that was used.

Each rule was used to classify the alignment of each simulated participant according to the simulated observed angles and simulated true angles. The observed classification was compared with the true classification to obtain the sensitivity and specificity for each classification rule. For a stability rule, the sensitivity is the probability that a truly stable alignment is classified as stable based on observed angles, and the specificity is the probability that a truly unstable alignment is classified as not stable based on observed angles. One minus specificity is the false-positive rate, that is, the probability that a truly unstable alignment is classified as stable. For an instability rule, the sensitivity is the probability that a truly unstable alignment is classified as unstable based on observed angles, and the specificity is the probability that a truly stable alignment is classified as not unstable based on observed angles. One minus specificity is the probability that a truly stable alignment is classified as unstable (false-positive probability).

SENSITIVITY ANALYSIS

In the sensitivity analysis, participants with patterns of change over time other than a constant increase or decrease were simulated, and other variables of the original simulations also were varied, such as baseline distribution of angles (mean, standard deviation, range, and form of distribution) and standard error of measurement, to determine how these factors affected results.

APPLICATION OF RESULTS TO AN OBSERVATIONAL STUDY

The Esotropia Treatment Study[7] was conducted by the Pediatric Eye Disease Investigator Group to investigate the stability of ocular alignment in childhood esotropia. Participants had follow-up visits every 6 weeks for 18 weeks for a total of 4 visits. Results for the classification rules based on 4 visits were used to select a rule for stability and a rule for instability, which then were used to classify participants in the Esotropia Treatment Study.

©2010 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 05/23/2013

Table 2. Sensitivity and 1 – Specificity for Stability Rules for a Fixed Number of Angle Measurements

Stability Rule[a]	Measurements, No.	Sensitivity, % (True-Positives)[b]	True Angle Change Between Successive Measurements, PD: 1 – Specificity,[c] % (False-Positives)											
			–10.0	–7.5	–5.0	–4.2	–2.5	–1.0	1.0	2.5	4.2	5.0	7.5	10.0
0 PD	2	34	6	13	22	23	29	32	33	32	27	25	16	9
	3	12	<1	<1	2	3	7	11	12	9	5	3	1	<1
	4	4	0	0	<1	<1	2	4	5	2	<1	<1	0	0
	5	2	0	0	0	<1	<1	1	2	1	<1	0	0	0
5 PD	2	80	27	45	62	67	75	79	80	77	70	67	52	35
	3	60	<1	4	20	27	47	58	59	49	34	26	7	1
	4	45	0	<1	3	5	23	40	41	25	9	4	<1	0
	5	33	0	0	<1	<1	8	26	28	10	1	<1	0	0
10 PD	2	97	64	80	90	92	95	96	97	96	93	91	81	69
	3	91	4	24	59	69	83	90	91	85	71	63	31	9
	4	86	0	1	19	33	67	82	83	67	38	26	3	<1
	5	79	0	0	2	8	45	73	74	45	12	5	<1	0
15 PD	2	>99	90	96	98	99	99	>99	>99	>99	99	99	96	91
	3	99	25	64	89	93	97	99	99	98	93	89	66	33
	4	98	<1	10	60	75	93	98	98	93	76	63	16	1
	5	97	0	<1	19	41	84	96	96	83	45	26	1	0

Abbreviation: PD, prism diopters.

[a] The rule applied is as follows: all measurements are within *x* PD, where *x* is the column value. True stability refers to no change in the true measurement. True instability refers to any change in the true measurement.

[b] Sensitivity is the probability that a truly stable alignment is classified as stable. For the purposes of this analysis, stable alignment was defined as zero change between successive measurements. For example, for a participant with 3 measurements with zero true change between measurements, the probability of a (true-positive) stable classification using the 5-PD rule for stability (classify as stable if all 3 measurements are within 5 PD) is 60%.

[c] 1 – Specificity is the probability that a truly unstable alignment is classified as stable, that is, it is the false-positive percentage for stability. For the purposes of this analysis, a truly unstable alignment was defined as having any nonzero amount of change between successive measurements. For example, for a participant with 3 measurements with a 4.2-PD true increase between each measurement, the probability of a (false-positive) stable classification using the 5-PD rule for stability (classify as stable if all 3 measurements are within 5 PD) is 34%.

RESULTS

SENSITIVITY AND SPECIFICITY OF RULES FOR A FIXED NUMBER OF MEASUREMENTS

The sensitivity and specificity of each of the classification rules for stability and instability for a fixed number of measurements are given in **Table 2** (stability) and **Table 3** (instability). For example, the sensitivity of a 0 PD rule for defining stability (all measured angles must be identical) applied to 4 measurements is 4%, whereas the sensitivity of the 5 PD rule (all measured angles are within 5 PD) applied to 4 measurements is 45%. In other words, because of measurement error, the 4 measured angles of a participant with an unchanging angle have only a 4% probability that they will be identical and a 45% probability that they will differ by no more than 5 PD (Table 2). Participants with a small change in the true angle of –2.5 or 2.5 PD between successive measurements across 4 measurements are slightly less likely to have measurements that do not differ (2% probability), whereas the probability that their measurements differ by 5 PD at most is 23% (for a –2.5-PD change) or 25% (for a 2.5-PD change). The slight asymmetry in specificity for an increase vs decrease in angle is due to several factors: (1) regression to the mean resulting from limiting the analysis to a range of baseline angles representing moderate esotropia, (2) higher precision of measurement for smaller angles (≤20 PD) compared with larger angles (>20 PD), and (3) rounding observed angles down to the nearest prism in the prism set. Nevertheless, these same effects would be expected in the clinical situation.

It is evident from Table 2 that the probability that a truly unstable alignment is misclassified as stable for any given rule depends on the amount of true angle change: the smaller the change, the more likely a participant's alignment will be misclassified as stable. Likewise, the larger the change in true angle, the more likely the alignment will be correctly classified as unstable for any given rule.

Applying the "all measurements within 5 PD" rule for stability to fewer than 4 measurements results in higher sensitivity for stability than with 4 measurements (60% for 3 measurements and 80% for 2 measurements) because there is less chance that a stable participant's measurements will differ by more than 5 PD due to measurement error when there are fewer measurements. However, the probability of a false-positive stable classification is substantially increased with fewer measurements.

If planning a study in which angle stability will be classified, the choice of number of measurements will depend on the false-positive and false-negative misclassification rates that can be tolerated. For example, for a target of 20% or less false-positive rates for angle changes corresponding to measurement error or greater (–4.2 or 4.2 PD per visit), one could take 3 measurements and apply the 0 PD rule, 4 measurements and apply the 5 PD rule, or 5 measurements and apply the 10 PD rule. The corresponding sensitivities for detecting stability are 12%, 45%, and 79%. With only 2 measurements, the false-positive rate for angle changes of –4.2 or 4.2 PD per visit

©2010 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 05/23/2013

Table 3. Sensitivity and 1 – Specificity for Instability Rules for a Fixed Number of Angle Measurements

Instability Rule[a]	Measurements, No.	1 – Specificity, % (False-Positives)[b]	True Angle Change Between Successive Measurements, PD: Sensitivity,[c] % (True-Positives)											
			–10.0	–7.5	–5.0	–4.2	–2.5	–1.0	1.0	2.5	4.2	5.0	7.5	10.0
10 PD	2	18	59	43	30	26	21	19	19	23	30	33	48	65
	3	37	99	89	66	59	44	37	38	50	66	74	93	99
	4	51	100	>99	92	85	63	52	56	74	91	96	>99	100
	5	62	100	100	>99	97	79	63	70	89	99	>99	100	100
15 PD	2	3	23	13	7	6	4	3	3	4	7	9	19	31
	3	7	85	54	26	20	12	8	8	14	29	37	69	91
	4	12	100	95	58	44	23	13	15	32	61	74	97	>99
	5	17	100	>99	90	75	37	20	24	54	88	95	>99	100
20 PD	2	<1	5	2	1	1	<1	<1	<1	1	1	1	4	9
	3	1	48	17	6	4	1	1	1	2	7	11	34	67
	4	1	98	68	20	13	4	2	2	7	24	37	84	99
	5	2	100	98	53	31	9	3	3	17	55	74	99	100
25 PD	2	0	1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	1
	3	<1	13	3	1	<1	<1	<1	<1	<1	1	1	9	32
	4	<1	81	25	3	2	<1	<1	<1	1	5	10	51	90
	5	<1	>99	82	14	6	1	<1	<1	2	19	37	93	>99

Abbreviation: PD, prism diopters.

[a] The rule applied is as follows: at least 2 measurements differing by *x* PD or more, where *x* is the column value. True stability refers to no change in the true measurement. True instability refers to any change in the true measurement.

[b] 1 – Specificity is the probability that a truly stable alignment is classified as unstable, that is, it is the false-positive percentage for instability. For the purposes of this analysis, stable alignment was defined as zero change between successive measurements. For example, for a participant with 4 measurements with zero true change between measurements, the probability of a (false-positive) unstable classification using the 15-PD rule for instability (classify as unstable if ≥2 measurements differ by ≥15 PD) is 12%.

[c] Sensitivity is the probability that a truly unstable alignment is classified as unstable. For the purposes of this analysis, a truly unstable alignment was defined as having any nonzero amount of change between successive measurements. For example, for a participant with 4 measurements with –5.0-PD true change between each measurement, the probability of a (true-positive) unstable classification using the 15-PD rule for instability (classify as unstable if ≥2 measurements differ by ≥15 PD) is 58%.

is greater than 20% regardless of what rule is used, so there is no acceptable rule for stability that is based on only 2 measurements.

For instability (Table 3), the rule that classifies a participant's alignment as unstable when a second measurement is observed that differs by 15 PD or more from any previous measurement has a false-positive rate of less than 20% when the number of measurements is 5 or less. When there are only 2 measurements, a rule that classifies that angle as unstable when the 2 measurements differ by 10 PD or more has a false-positive rate of 18%. For classifying instability, the false-positive rate decreases as the number of measurements decreases because the likelihood that 1 measurement will differ by any given amount (or more) from other measurements due to measurement error alone is reduced when there are fewer measurements.

RULES FOR PROSPECTIVE CLASSIFICATION OF STABILITY AND INSTABILITY

Of interest to the clinician is the ability to classify stability during a series of return visits. A set of rules for classification of stability was developed by selecting the pragmatic rule for each possible choice of stability cutoff value (0, 5, 10, or 15 PD) that had a 1 – specificity (false-positive rate) no greater than 20% for per-visit angle changes that met or exceeded measurement error and was based on the smallest number of measurements meeting this criterion. We also considered combinations of the pragmatic rules that would be expected to meet this criterion. This resulted in the rules for prospective stability classification given in Table 1.

The sensitivity and specificity for stability classification with prospective follow-up depended on the number of visits that one was willing to continue following participants who have not yet met the criteria for stability (data not shown). At least 3 visits were required. As the number of visits was increased, the sensitivity increased but the specificity decreased because the probability that a participant with changing angle would meet stability criteria by chance increased as the number of visits increased. Chances of misclassification of a participant with a large angle change (more than measurement error) remained low as the number of visits was increased, but the chances of misclassification of a participant with an angle change smaller than measurement error were greatly increased as the number of measurements increased. There was not a lot of difference in rule performance for angle changes that were much larger or smaller than measurement error, but for angle changes in the vicinity of measurement error, rule 2 (classify as stable if the last 4 consecutive measurements are within 5 PD) performed slightly better than the other proposed rules. Rules meeting the criterion of 20% or less false-positive rates that are based on more than 4 measurements would be expected to perform somewhat better yet.

SENSITIVITY ANALYSES

Neither the stability rule nor the instability rule was sensitive to the population mean angle, the standard devia-

©2010 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 05/23/2013

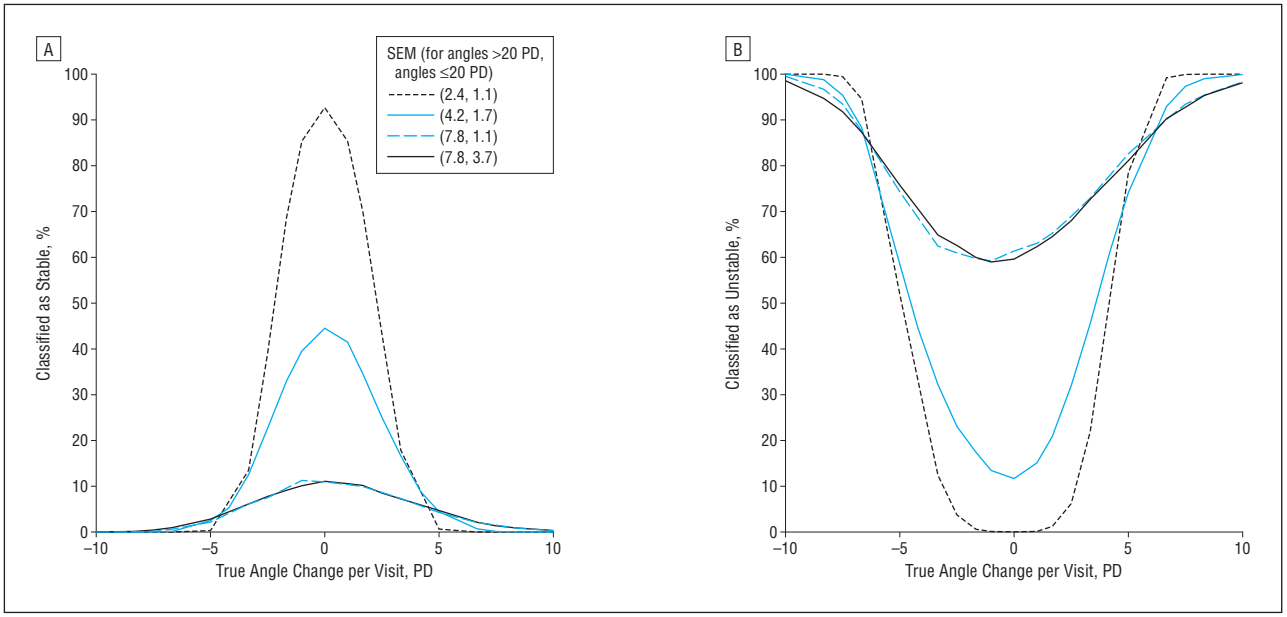

Figure 1. Effect of measurement error on the probability of stable (A) and unstable (B) classification for the rule "all 4 measurements are within 5 prism diopters (PD) (see Appendix)" as a function of true angle change. The base assumption (standard error of measurement [SEM] was 4.2 for angles >20 PD and 1.7 for angles ≤20 PD) is represented by the solid black line. The broken lines show the effect of varying the base assumption.

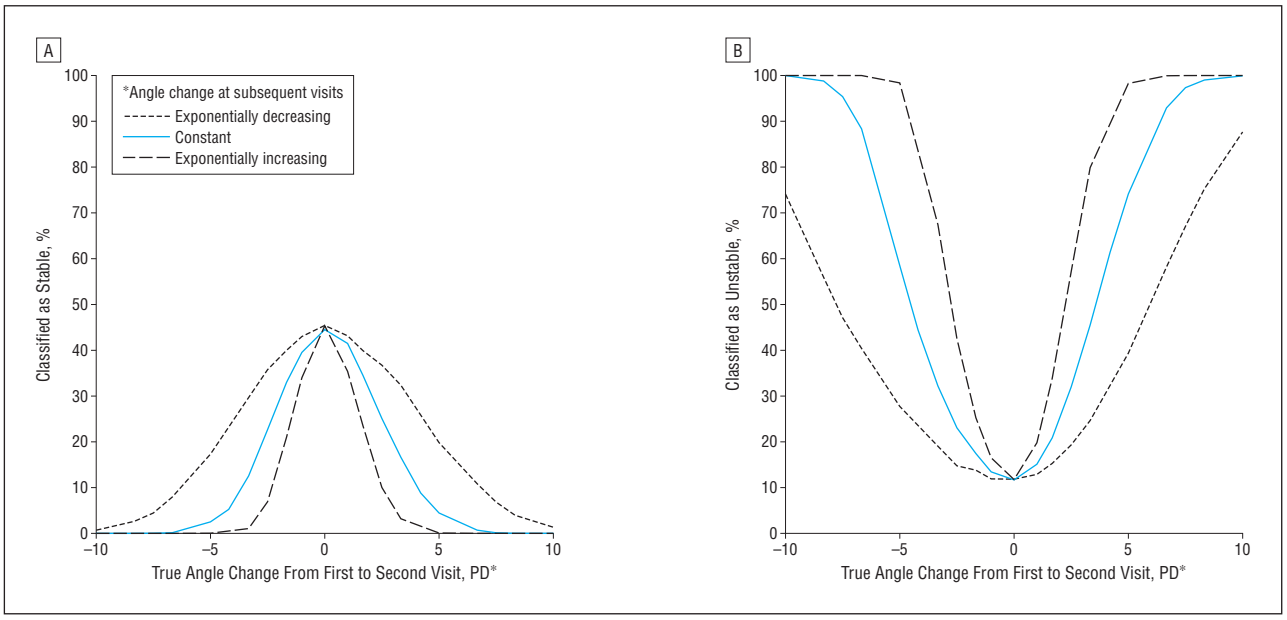

Figure 2. Effect of pattern of angle change on the probability of stable (A) and unstable (B) classification for the rule "at least 2 measurements differing by 15 PD (see Appendix) or more" as a function of true angle change. The base assumption (constant rate of angle change) is represented by the solid black line. The broken lines show the effect of varying the base assumption.

tion of the population angles, the shape of the population angle distribution, or the range of the population angle distribution (data not shown). There was a very slight effect of changing the population mean or range that was due to a different standard error of measurement for smaller angles (≤20 PD) as opposed to large angles (>20 PD) and regression to the mean. For example, when the mean angle was smaller, a greater proportion of the population was measured with the smaller measurement error found for smaller angles, and this resulted in better specificity for classifying negative angle changes. However, there also was regression to the mean in the negative direction in a greater proportion of the population, which resulted in worse specificity for classifying positive angle changes. Varying standard error of measurement, as expected, had a large effect on classification: the smaller the measurement error, the more accurate the classification (**Figure 1**). Likewise, varying the pattern of angle change affected the accuracy of classification (**Figure 2**). Participants whose angle was changing but the amount of change was getting smaller over time were harder to classify correctly (as unstable) than those with constant angle

©2010 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 05/23/2013

change, who were, in turn, harder to classify than those whose angle was changing but the amount of change was getting larger over time. Similar results were obtained for the prospective classification rules and for the other fixed rules, regardless of the number of measurements, that is, the results were sensitive to only the standard error of measurement and the pattern of angle change.

APPLICATION OF RESULTS TO THE ESOTROPIA TREATMENT STUDY

For classifying alignment of participants in the Esotropia Treatment Study,[7] it was desired to maximize sensitivity while keeping the percentage of false-positives for stability and instability lower than 20% for per-visit angle changes exceeding measurement error (4.2 PD per visit). The "all measurements within 5 PD" rule was adopted for stability, despite its low sensitivity (45%) because less stringent rules did not meet the specified false-positive criterion. Likewise, the rule for instability of "at least 2 measures differing by 15 PD or more" led to the highest sensitivities while meeting the target of less than 20% false-positive results (ie, 12%). Using these rules, 46% of 59 participants with infantile esotropia were classified as having unstable ocular alignment, and 20% were classified as having stable alignment. For 60 participants with acquired nonaccommodative esotropia, the corresponding figures were 22% unstable and 37% stable, and for 41 participants with acquired partially accommodative esotropia, the figures were 15% unstable and 39% stable.[7]

COMMENT

We modeled "changes in the observed angle of misalignment," using estimates of measurement error based on a test-retest study, to test potential rules for stability and instability in childhood esotropia. For a series of 4 measurements, such as used in the Esotropia Treatment Study,[7] the rules that optimized sensitivity and specificity were "all measurements within 5 PD" as the rule for stability and "at least 2 measures differing by 15 PD or more" as the rule for instability. For the more common clinical scenario of the performance of an undefined number of prospective measurements, we found that a minimum of 3 identical measurements or 4 measurements within 5 PD would be needed to confirm stability.

In the present modeling, we also noted that chances for prospective misclassification of alignment of a participant with changing (unstable) alignment as stable are high for modest angle changes (≤5 PD per visit), and so clinically it would be reasonable to space visits far enough apart to give changes time to manifest. The optimum timing of such visits needs to be the subject of a future study.

In considering the practical application of these results, one approach to account for measurement error in strabismus measurements might be to make multiple measurements during the same office visit. Such an approach would require the clinician to make each measurement independently, accepting that the previous measurement might have been unrepresentative of the true misalignment. Ignoring the previous measurement and approaching the next measurement independently a few minutes later is difficult if not impossible. Acquiring multiple unbiased measurements during the same office visit may require a more automated means of measuring the magnitude of strabismus, such as eye tracker technology.[8]

One potential weakness of this study was highlighted by the sensitivity analysis; the sensitivity and specificity of the proposed classification rules were highly dependent on the magnitude of measurement error. Nevertheless, we believe that the present estimates of prism and alternate cover test measurement error are reasonable because they were derived from a prospective test-retest study specifically designed for the purpose of estimating measurement error.[6] The sensitivity and specificity of the proposed classification rules remained high even when the population mean, standard deviation, range, and shape of the angle distribution were varied. In conclusion, clinically useful definitions of stability and instability in childhood esotropia have been derived using estimates of actual measurement error and may be reasonable to apply to clinical practice and future strabismus studies.

Submitted for Publication: December 22, 2009; final revision received April 8, 2010; accepted April 13, 2010.
Correspondence: B. Michele Melia, ScM, Jaeb Center for Health Research, 15310 Amberly Dr, Ste 350, Tampa, FL 33647 (pedig@jaeb.org).
Pediatric Eye Disease Investigator Group: A full list of the members is available at http://www.ncbi.nlm.nih.gov/pubmed/18973948.
Financial Disclosure: None reported.
Funding/Support: This study was supported by grants EY011751 and EY015799 from the National Eye Institute of the National Institutes of Health, Department of Health and Human Services.
Previous Presentation: This study was presented in part at the annual meeting of the Association for Research in Vision and Ophthalmology; April 28, 2008; Fort Lauderdale, Florida.
Online-Only Material: The eAppendix is available at http://www.archophthalmol.com.

REFERENCES

1. Birch EE, Felius J, Stager DR Sr, Weakley DR Jr, Bosworth RG. Pre-operative stability of infantile esotropia and post-operative outcome. *Am J Ophthalmol*. 2004;138(6):1003-1009.
2. Kitzmann AS, Mohney BG, Diehl NN. Progressive increase in the angle of deviation in acquired nonaccommodative esotropia of childhood. *J AAPOS*. 2003;7(5):349-353.
3. Lueder GT, Galli ML. Effect of preoperative stability of alignment on outcome of strabismus surgery for infantile esotropia. *J AAPOS*. 2008;12(1):66-68.
4. Ing MR. Progressive increase in the angle of deviation in congenital esotropia. *Trans Am Ophthalmol Soc*. 1994;92:117-125, discussion 126-131.
5. Beck RW, Maguire MG, Bressler NM, Glassman AR, Lindblad AS, Ferris FL. Visual acuity as an outcome measure in clinical trials of retinal diseases. *Ophthalmology*. 2007;114(10):1804-1809.
6. Pediatric Eye Disease Investigator Group. Interobserver reliability of the prism and alternate cover test in children with esotropia. *Arch Ophthalmol*. 2009;127(1):59-65.
7. Christiansen SP, Chandler DL, Holmes JM, et al; Pediatric Eye Disease Investigator Group. Instability of ocular alignment in childhood esotropia. *Ophthalmology*. 2008;115(12):2266-2274, e4.
8. Leske DA, Hatt SR, Laird PW, Arrington KF, Holmes JM. Infrared eyetracker methods for measuring strabismus [abstract]. *Invest Ophthalmol Vis Sci*. 2007;48:E-Abstract 900. http://www.arringtonresearch.com/kfa/abstract/Infrared%20Eyetracker%20MethodsforMeasuringStrabismus.pdf. Accessed August 28, 2009.

©2010 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 05/23/2013