

## QUALITY ASSESSMENT PROGRAM IN HISTOPATHOLOGY: A PILOT STUDY FROM MAHARASHTRA

SHOBHA AGASHE, ALKA GOSAVI, RAKHEE JAGDALE\*, JAYDEEP POL\*,

*Department of Pathology, Lecturer in Pathology\*, Govt. Medical College, Miraj, Maharashtra, India*

### ABSTRACT

Histopathology reports are important quality assurance tools and evaluation of pathological diagnoses described in them is an integral part of total quality control and quality improvement program. We describe a program based on slide circulation which was aimed at both continuing education to upgrade knowledge and proficiency testing of histopathologists. The performance of the participating pathologists was analyzed and the degree of agreement was also studied. The results showed improvement indicated by rising level of performance in 35.3% of consistent participants and increasing trend in the average score. The degree of agreement was comparatively low (65.29%). The practicability of this program and its acceptability as an EQAS was also investigated.

**Key Words:** Quality Assurance (QA) in histopathology, proficiency testing, External Quality Assessment Scheme (EQAS).

### INTRODUCTION

Quality may be defined as a measure of excellence and is an important aspect of any laboratory. A laboratory must give consistently correct results as the patient is benefited by this. It is also expected that similar results are obtained on the same material at different centers. Various quality control and quality improvement programs have been advocated to measure the efficiency of techniques and proficiency of pathologists<sup>1-14</sup>. In the literature, the quality control programs in clinical pathology and hematology are now widely accepted throughout the world; but histopathologists have been slow to adopt this concept. The reluctance is usually ascribed more to the belief that evaluation of histopathological diagnosis is impracticable as it deals with opinions and is largely subjective. The diagnostic opinions which are often predictions of behavior of a morbid process and verified only by future events, may seem objectively untestable. Nevertheless, the likelihood of correct diagnosis generally reflects the professional training, experience, good judgement and competence of interpreter and hence evaluation in histopathology should be possible if meaningful and reliable methods are developed.

Quality assurance in histopathology is desirable and includes the evaluation of both processes and outcomes<sup>7</sup>. The processes are concerned with the receipt, description and processing of specimens while outcomes include pathological diagnoses and the reports that describe them. It also consists of timely clear communication and schemes for quality

assessment in histopathology which have been tried in the past. For any quality assurance program it is desirable to maintain a document that describes it and should include the journalistic elements of "Who, what, when and how"<sup>7</sup>. In this article, we report the result of a study which consists of circulation of slides at state level and evaluation of performance of the participating pathologists. The study is similar to Australian study<sup>1</sup> and West of Scotland study and fulfills the design considerations recommended by National EQA forum U.K<sup>6</sup>.

### MATERIAL AND METHODS

Maharashtra state chapter of Indian Association of Pathologists and Microbiologists was conducting slide circulation project for last several years. It was an important avenue of continuing educational program. But in spite of good histopathological material, the participation was not active and the approach was quite casual. To make this project more attractive scoring system was added and some of the circulated slides having diagnostic controversies were identified and these were discussed in a histopathology slide seminar at the Annual Conference. The results of this exercise were declared and the awards were given to the top scorers during this annual meet. This venture aroused great interest in the participating pathologists and more efforts were taken to reach the top. The team led by one of the authors (S Agashe) participated in these programs and received first rank with an award of 'Excellence in Histopathology' twice. Later on in 1997, she was entrusted by the state chapter authorities to conduct

this program for next 3 years. This challenging work was accepted willingly and enthusiastically by her.

We conducted this program for 3 years. It consisted of circulation of histopathology slides along with evaluation of the diagnoses offered by participating pathologists. Each year, all the academic centers and private pathologists, keeping academic interest from Maharashtra state were requested to contribute 3 to 4 interesting histopathologic lesions from their routine work. This was the first communication in which detail instructions were given for selecting the lesion and sending the slides. They were requested to prepare and send 25-30 slides from the same tissue block each slide showing the representative lesion. The diagnosis must be possible by studying one or two provided sections. Lesions requiring many sections to make a diagnosis or which need further studies like immunohistochemistry were included for the circulation, but they were not considered for evaluation. The diagnosis of these cases was not revealed until such cases were presented and discussed by the contributors, in the slide seminar during the Annual Conference. The contributors were also asked to send the slides along with brief clinical details, relevant investigations, gross findings but no diagnosis. All the slides were collected by the convenor. The slides were coded. Names of the contributors were not disclosed. This was to have complete unbiased opinion on the slides which may be associated with the name of the specialized center or developing institute. The slide sets along with clinical information were distributed to all the contributors. This was the second communication which invited list of diagnoses on these slides in a span of 2 months. Along with this, contributors were requested to send the drafted report (microscopic findings) with the diagnosis of the cases, which they had contributed.

Most of the contributing centers participated in the evaluation program but there were few dropouts. There were 30-33 contributors and 23-24 participants every year. Only 17 centers were consistent in sending the list of diagnosis for 3 consecutive years. After receiving list of diagnosis from all the participants for assessment, a list of contributors' diagnosis was sent to all.

The assessment was done by the convenor using contributors' diagnosis as the reference diagnosis. The answer sheets submitted by the participants were also coded. During assessment,

the degree of agreement on all the slides was studied which was denoted by Facility Index (F. I.). Facility Index points out the easiness of the slide and is a good indicator of agreement. e.g. a slide with 100% F.I. indicates disagreement. Slides with low F.I. i.e. <30% were reopened and the contributors were requested to present and discuss these cases in a slide seminar during the Annual Conference.

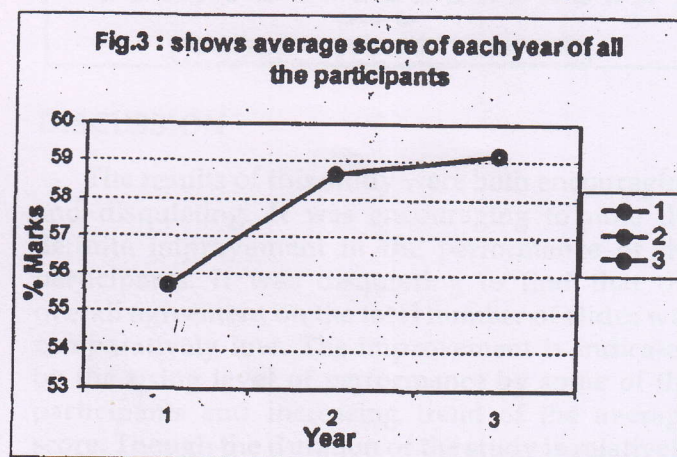
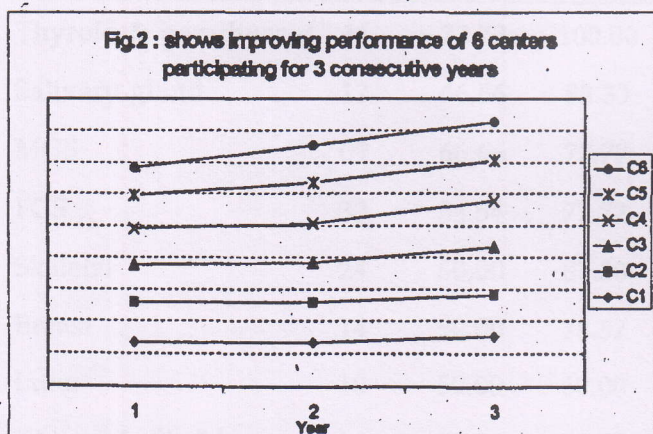
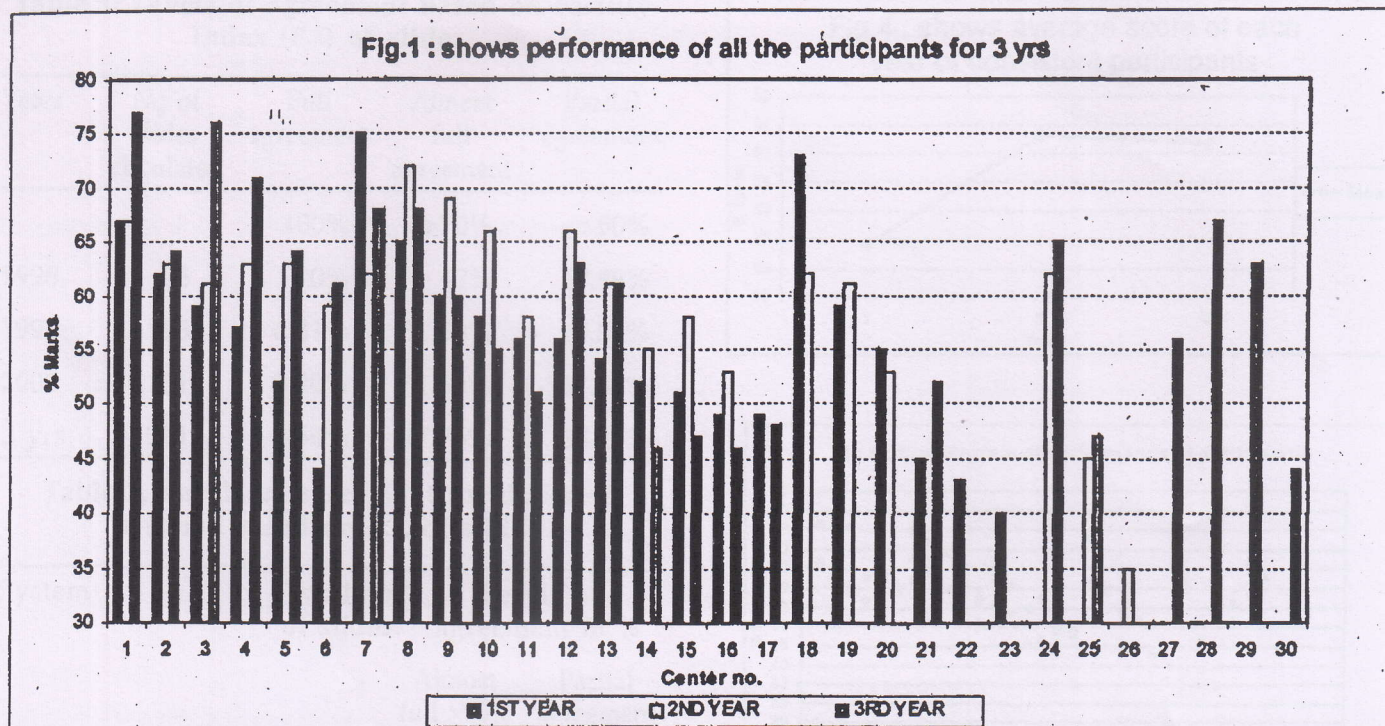
It was very interesting to note an almost unanimous opinion by majority of the participants on certain slides, which was different from that of the contributor. Obviously such slides had low F.I. These slides were studied and an Agreement Index (A.I.) was calculated. A. I. means the percentage of participants agreeing upon one diagnosis which is different from that given by the contributor. In other words A. I. indicates consensus opinion.

While calculating the marks of a particular center the slides contributed by the center and the slides for discussion were deleted and then mark sheet of each participant was prepared. In this way, the performance of the participants was judged and the degree of agreement on all slides was analyzed. The results were declared during the Annual Conference and certificates and awards were given to the top scorers. At the end of each year, a file was sent to all the contributors, which contained drafted report of each case along with the diagnosis, facility index of each case with graphical presentation, A.I. when it was > 70% suggesting consensus diagnosis score of the said center and graphical presentation of the performance of all the participants. This was to make them able to assess their own performance in relation to average but keeping the secrecy of the marks obtained by other participants.

## OBSERVATIONS AND RESULTS

The results of this pilot study consist of analysis of performance of the participants and the degree of agreement on the slides.

*Performance of participants:* The number of centers participating for assessment varied each year which was 23 for the first and second year and 24 for the third year. As the centers participated variably, the total number of centers participating for three years were 30. Only 17 centers were consistent in participating for all the three years (Fig 1). This also shows that there were few dropouts who showed either poor performance (center no 21, 22 and 23) or deterioration in performance (center no 18, 20). There were 4 new participants in the third year indicating the



increasing acceptability and popularity of the program. It was interesting to note that out of the consistent 17 centers, 6 (35.3%) showed improvement in their performance in successive years (Fig 2). The average score of each year was calculated considering all the participants in that year, which showed an increasing trend (Fig 3). The average score of each year of the consistent participants was also calculated which showed a notable increase in the second year and a negligible decline in the third year (Fig 4). A scattergram of performance (Fig.5) of the participants was plotted. The centers with poor performance can be easily identified as they fall outside the main cluster.

**Agreement on slides:** The degree of agreement

based on facility index (F.I.) of slides was also analyzed. A high value of F.I. indicates good agreement. The slides were studied to look for the level of agreement (Table 1). Though, there were very few slides with full agreement (i.e. 100% F.I.), the number of such cases increased each year. The other levels of agreement i.e. almost full and partial agreement also increased in the second year and remained almost same in the third year. The overall agreement achieved on 291 slides was 65.29%, which means >50% of participants agreed on 65.29% slides. Similar levels were calculated for benign and malignant tumors which revealed almost equal agreement viz. 63.63% and 64.22% respectively. Table no 2 shows the extent of agreement reached on the slides from different tissues. The greatest

**Table 1: Levels of agreement based on Facility Index (F.I) of slides.**

| Years | No of slides circulated | Full agreement | Almost full agreement | Partial agreement |
|-------|-------------------------|----------------|-----------------------|-------------------|
|       |                         | 100%           | >70%                  | >50%              |
| 1998  | 95                      | 0%             | 40%                   | 57.89%            |
| 1999  | 95                      | 6.31%          | 51.50%                | 70.53%            |
| 2000  | 101                     | 9.90%          | 51.50%                | 67.33%            |
| 3 yrs | 291                     | 5.50%          | 47.76%                | 65.29%            |

**Table 2: Levels of agreement for different systems based on F.I. of slides**

| System                | Total no of slides | Levels of agreement in % |                        |
|-----------------------|--------------------|--------------------------|------------------------|
|                       |                    | Almost full >70%         | Partial agreement >50% |
| Thyroid & parathyroid | 11                 | 72.72                    | 100.00                 |
| Salivary gland        | 12                 | 66.66                    | 83.33                  |
| MGS                   | 09                 | 66.66                    | 77.77                  |
| FGS                   | 33                 | 54.54                    | 72.72                  |
| Skeletal              | 24                 | 50.00                    | 87.50                  |
| Breast                | 14                 | 50.00                    | 78.57                  |
| Lung                  | 10                 | 50.00                    | 60.00                  |
| Kidney & Bladder      | 16                 | 50.00                    | 68.75                  |
| Skin                  | 31                 | 48.38                    | 61.29                  |
| Nervous system        | 17                 | 47.05                    | 76.47                  |
| GIT & Liver           | 46                 | 45.65                    | 60.86                  |
| Infective etiology    | 23                 | 47.45                    | 65.21                  |
| Lymph node            | 22                 | 45.45                    | 68.18                  |
| Soft tissue           | 22                 | 22.72                    | 36.36                  |

degree of agreement was achieved on lesions of thyroid, salivary gland and male genital system while least on lymph node, GIT and soft tissue tumors. This study is an ongoing process and we expect database for analysis in future.

**Fig.4 : shows average score of each year of consistent participants**

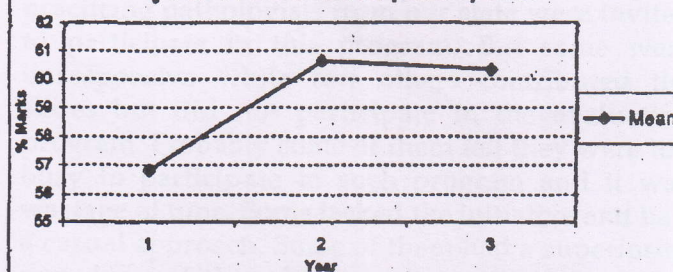
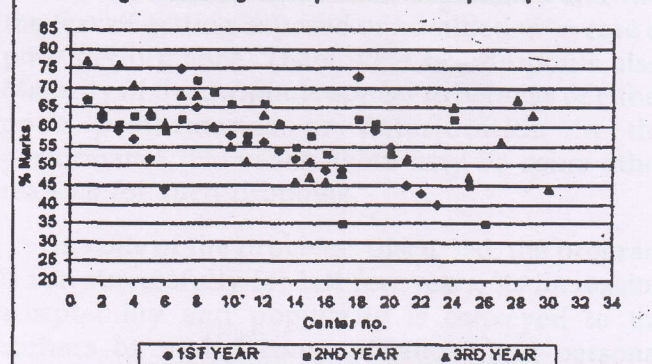


**Fig.5 : Scattergram of performance of participants**



## DISCUSSION

The results of this study were both encouraging and disquieting. It was encouraging to note the definite improvement in the performance of the participants. It was disquieting to find that the overall agreement on the total number of slides was comparatively low. The improvement is indicated by the rising level of performance by some of the participants and increasing trend of the average score. Though the duration of the study is relatively short to draw any inference, we hope that continuation of such a program will increase awareness about the proficiency and will definitely improve the quality of histopathologists in future.

In histopathology when a group of pathologists agree on a diagnosis, it is likely to be the correct one. Determination of accuracy is practically a matter of consensus. Substantial agreement among pathologists though not an absolute measure of accuracy is effectively the best available measure of accuracy. Various statistical methods have been used for measuring levels of agreement<sup>15,16</sup>. We have used a simple method for calculating levels of agreement and have emphasized on analyzing various reasons for disagreement. We agree with Sherwood and Hunt who in their concluding

remarks have said that studies based on statistical analysis amply show that pathologists differ but such study would no longer be an EQAS. Our study reveals low level of agreement as compared to others (Owen -77%, Penner-78%, Sherwood\_77%, Husain-87%). The reason for this may be that we have assessed the slides assuming contributor's diagnosis as the correct diagnosis. while others have considered consensus opinion or the expert panel opinion as the final diagnosis. another reason for the discrepancy may be that the slides were not identified as those of the general pathologists' interest and those of the specialist pathologists' interest. The concept of specialist pathologist is new to Indian scenario, except for few centers.

Most of the pathologist are general pathologists dealing with all the branches of histopathology. e.g. a general pathologist is unlikely to be conversant with skin adnexal tumors which on the other hand is a day to day affair for a dermatopathologist. The pathologists participating in this programs were from metros as well as from general hospitals in small towns. This might have affected the agreement level.

While analyzing the causes of low agreement level it was also pointed out that at times the diagnosis offered by the contributor was wrong or the terminology used was improper or vague. This was indicted by disagreement on such slides by all the participants. This may be a fallacy of this study but we thought it was necessary to increase the palatability of the program, especially during preliminary stages of introduction of any EQAS in histopathology in this region.

It was also observed that in spite of proper instructions to the contributing pathologists, some of the slides included for assessment were very rare or difficult lesions and needed further studies to arrive at a diagnosis. Naturally the agreement level for these slides was low. Some of such controversial slides were reopened and were reviewed and discussed at the annual meet. The low agreement level also reflects on the performance of the participants. We therefore like to suggest that a panel of experts should screen the slides before they are accepted for circulation. Contributor's diagnosis should not be taken as a reference diagnosis for the assessment of the participants. Instead, participants should be evaluated against expert panel opinion or consensus opinion meaning thereby the diagnosis offered by more than 50% of the participants. Though all this may be the wishful thinking, it gives

us the nightmare of unpopularity of the program

Every year, all the academic institutes and practicing pathologists from our state were invited to participate in this program. But some were unresponsive while few others contributed the slides but did not participate in the evaluation program. Probably some of them felt they were too busy to participate in such program and it was a waste of time. Some lacked the initiative and had a casual approach. Some of them had a superiority complex and thought their diagnosis is final and cannot be evaluated by anyone else. At the other end, there were centers from smaller towns who lacked confidence more than experience and had the fear of getting exposed or humiliation in case of poor performance. There were few dropouts also. Majority of the dropouts appear to be because of either poor performance or deterioration in the performance. However, there may be some other reasons for such dropouts.

In spite of the problems discussed, the program is run successfully for last few years. Its increasing acceptability and popularity is conveyed to the authors by verbal communications and personal letters by various pathologists. It is also indicated by newer entries.

The project outlined in this paper involves a slide club run at a state level along with evaluation of participants. Our scheme does not directly evaluate the quality of performance as it is actually practiced and as it directly affects patients care, as advocated by Penner. The use of randomly selected material from patient file is advised by him so that routine diagnostic material is tested. However, this does not arouse interest in the participants and may not be of any educational value. The organizers of West of Scotland scheme are of the opinion that the response is better to material that the participants consider both of interest and having educational content. We are of the same opinion and feel that the results of our study will indirectly reflect on day to day practice.

## CONCLUSIONS

Our scheme fulfils most of the design considerations recommended by the National EQA Forum, U.K, regarding the selection of material, qualification of organizer, number of participants and reporting of results. The project is a good educational exercise for consultants and trainee pathologists fulfilling the aim of continuing medical education. It allows the histopathologist to monitor

his/her performance in relation to the average. It is possible to identify the bad performer so that some remedial action can be taken. It also allows the participant pathologist to compare his reporting style and diagnosis with his colleagues. The discussions on controversial slides also help in continuing education.

In conclusion, we feel that this type of project will be a dual track educational and quality assurance program and with few modifications can be adopted as a statewide EQAS in histopathology, also supplementing the internal audit.

#### References

1. Penner D. W. Quality control and quality evaluation in histopathology and cytology. *Pathol. Annu*, 1973; 8: 1-19.
2. Evans D.M.D., Shelly G, Cleary B, Baldwin Y. Observer variation and quality control of cytodiagnosis, *J. Clin. Pathol*, 1974; 27: 945-50.
3. F. A. Langley. Quality control in histopathology and diagnostic cytology, *Histopathology*, 1978; 2: 3-18.
4. Macartney J.C, Henson D. E, Codling B.W. Quality assurance in anatomic pathology, *Am. J. Clin. Pathol*, 1981; 75 (Suppl): 467-75.
5. Whitehead T.P, Woodford F.P. External quality assessment of clinical laboratories in the United Kingdom, *J. Clin. Pathol*, 1981; 34: 947-57.
6. Lee F.D, Burnett R.A. Quality assurance in histopathology, *J. Pathol*, 1987; 152: 247-51.
7. Henry Travers. Quality Assurance in Anatomic Pathology, *Laboratory Medicine*, 1989; 85-92.
8. Cowan D. F Quality assurance in anatomic pathology An information system approach, *Arch. Pathol. Lab Med*, 1990; 114: 129-34.
9. Rickert. Quality Assurance Goals in Surgical Pathology, *Arch. Path. Lab. Med*, 1990; 114: 1157-62.
10. QC/QA committee of the Association of Directors of Anatomic and Surgical Pathology. Recommendations on Quality control and Quality assurance in Anatomic Pathology, *Am. J. Surg. Pathol*, 1991; 15: 1007-09.
11. Peter P. Anthony, Roderick N. M. MacSween. Recent advances in Histopathology, No. 16 New York Churchill Livingstone, 1994; 293-309.
12. Sherwood A.J, Hunt A.C. An external quality assessment scheme in histopathology, *J. Clin. Pathol* 1984; 37: 409-14.
13. Husain OAN, Butler E B, Woodford F.P. Combined external quality assessment of cytology and histology opinions. A pilot scheme for a cluster of laboratories *J. Clin. Pathol*, 1984; 37: 993-1003.
14. Owen D. A, Tighe J. R. Quality evaluation in histopathology, *Br. Med. J*, 1975; 1: 149-50.
15. Holman C. D J, James I R, Heenan P J et al. An improved method of analysis of observer variation between pathologists, *Histopathology*, 1982; 6: 581-89.
16. Silcocks PBS Measuring repeatability and validity of histological diagnosis- a brief review with some practical examples, *J. Clin. Pathol*, 1983; 36: 1269-75.

---

Reprint Address : Dr. Alka Gosavi. M.D (Pathology)  
798, Behind Dr. Limaye's Eye Hospital,  
Vishrambag, Sangli. 416415

D/R - 17-08-01

D/A - 27-12-01