Clinics in Oncology

Research Article
Published: 05 Dec, 2016

Automatisation of the Gleason Score, The End of the Dilemma? A Proposal to Increase Consensus

*J Varinot[1], H Alshehhe[1], A Furudoi[1], F Soussaline[2], M Soussaline[2], R Renard-Penna[3], P Mozer[4], O Cussenot[5] and E Compérat[1,5]**

[1]Department of Pathology, Hôpital La-Pitié Salpêtrière, UPMC Paris VI, France

[2]Imstar, France

[3]Department of Radiology, Hôpital La-Pitié Salpêtrière, UPMC Paris VI, France

[4]Department of Urology, Hôpital La-Pitié Salpêtrière, UPMC Paris VI, France

[5]CeRePP, UPMC Paris VI, France

Abstract

Introduction: The updated Gleason score (GS) (WHO classification 2016) decreases interobserver variability, but discordance still exists. The aim was to apply an operator independent method by a computer system (IMSTAR) to give a GS via automated image analysis.

Material and Methods: Twenty-six prostate biopsies (PB) were evaluated by three pathologists. GS was reported, new slides of the same PB were double stained by immunofluorescence (Annexin-3, a marker of low GG and normal tissue, and p504s, a marker of Pca employed in routine) . The slides were scored using Pathfinder™ reader analyzer, for quantitative digital pathology automated scanning of the PB. Detection of different fluorescent signals allowed with a specific algorithm to give a GS by the computer.

Results: The interobserver consensus was 96%. When comparing pathologists and the computer analysis, we found a 100 % agreement in 13 cases (50%). In 8 (31%) cases, GS assigned by pathologists was lower than that assigned by the computer. In 5 cases (19%) pathologists attributed a higher GS than the computer. Discordance never exceeded 1 GG. Major discrepancy existed between the GS 6 and 7.

Conclusion: This study shows feasibility of GS standardization. Automated image analysis seems to be a promising operator independent technique to standardize GS and provide more homogeneous grading for PCa.

Keywords: Prostate cancer; Gleason grade; prostate biopsy; Quantitative digital pathology

OPEN ACCESS

***Correspondence:**
Eva Compérat, Department of Pathology, Hôpital La-Pitié Salpêtrière, UPMC Paris VI, France, Tel:0033142177788; Fax: 0033142177777;
E-mail: eva.comperat@aphp.fr

Received Date: *26 Oct 2016*
Accepted Date*: 10 Nov 2016*
Published Date*: 05 Dec 2016*

Citation:
Varinot J, Alshehhe H, Furudoi A, Soussaline F, Soussaline M, Renard-Penna R, et al. Automatisation of the Gleason Score, The End of the Dilemma? A Proposal to Increase Consensus. Clin Oncol. 2016; 1: 1159.

Introduction

Prostate Cancer (PCa) is recognized as one of the major medical problems facing the male population. In France, 71220 new cases of PCa were diagnosed in 2011 and it is considered the third leading cause of cancer death among men and accounts for approximately 8700 deaths per year [1].

The histopathological analysis of the prostate biopsy is the key step in diagnosing PCa. The Gleason score (GS) defining the differentiation of the PCa is one of the most important prognostic parameters in handling these patients and one of the strongest predictor of outcome.

During the last decade considerable changes in the practical application of Gleason grading (GG) have been made. Pathologists have made efforts towards standardization of GG. In 2005 the ISUP (International Society of Urological Pathology) revised the GG system, which led to major changes especially on prostate biopsy material [2]. A new consensus meeting was held in 2014, as new handling in daily practice has come up, especially GS 2-5 is not given any more. GS 6 has become the lowest score. The aim of these changes was to obtain a better correlation between biopsy and radical prostatectomy, less inter-observer variability, more heterogeneous treatment groups, but also to better predict biochemical recurrence.

Briefly, major changes in 2005 included: to report any component of higher grade, to omit GS 1+1=2, not to give 2+2=4 on biopsy material, to exclude GG 3 with individual cells, to consider cribriform patterns as GG4, to redefine grading variants for some histologic subtypes, to report any high grade tumor, to add tertiary patterns and to assign individual GS for every biopsy. The most important changes concerned GG 3 and 4, with an important limitation of GG3 and an increasing importance of the definition of GG4.

The new grading system was adopted by pathologists, as it also allows a more uniform interpretation between pathologists.

Nevertheless, urologists had to handle the impact of the upgrading and adopt the new system in their nomograms such as the D'Amico nomogram, which is essentially based on the GS. Although better uniformity of prostate biopsies was achieved with the new recommendations, there still remains a part of "individual interpretation". Even among experts in uropathology, the inter observer variability still remains around 30-40%. The major problem is the distinction between GG 3 and 4 [3].

Pierorazio and the Johns Hopkins team suggested GG groups for a more accurate system, with group I-V (GS 6=group I, GS 3+4= group II, GS 4+3 =group III, GS 4+4= group IV and GS 9-10= group V). This suggestion has been accepted and was adopted in the new World Health Organisation 2016 [4].

For all these reasons an automation of the GS would be highly appreciated by the community. The aim was to perform the GS using computer-assisted quantitative analysis, in order to achieve of a more consistent/reproducible scoring system. We compared the GS, and the GG groups diagnosed by three independent pathologists, used to work together and compared their results with the computer system's interpretation. Therefore, we double stained with an immunofluorescence technique prostate biopsies with two markers: Annexin 3 (A3) and p504s.

Annexin A3 (ANXA3) is member of a family of calcium-binding protein, implicated in several functions of the cell, such as apoptosis, inflammation and especially specific immune responses. It is expressed in healthy epithelial cells, exhibits strong staining in precancerous prostatic intraepithelial neoplasia (PIN), and is relatively less abundant in individual tumour cells of increasing GS, despite exhibiting higher overall tissue abundance in tumours. ANXA3 staining is predominantly cytoplasmic [5].

Recent studies showed that Annexin A3 (ANXA3) is a complementary marker of PSA and significantly associated with low grade PCa, GS 6 (3+3). Recent studies could also demonstrate that ANXA3 is a promising tissue marker and could provide prediction on prognosis in the individual patient.

P504S, also known as alpha-methylacyl-CoA racemase, recently identified by cDNA subtraction and microarray technology, serves as a specific marker because it has been demonstrated to be highly expressed in PCa, but not in benign prostatic glands. It is used in daily routine and cytoplasmic staining is an argument in favour of PCa [6].

After superposition of both markers und an immunofluorescence microscope, the level of orange-red expression should be correlated with the GS of the slide by computer analysis.

Materials and Methods

Specimen preparation

In this preliminary study, 26 standard prostate biopsies obtained from 21 patients, after MRI and target bhiopsies, with the diagnosis of PCa, were stained with standard Hematein-Eosine-Soffran(HES) staining in our pathology department (La Pitié-Salpêtrière Hospital). The biopsies were evaluated by one senior and two junior pathologists according to the recommendation of WHO 2016. GG was evaluated on each biopsy, in case of discrepancy, a consensus reading was performed. No biopsy with a tertiary pattern was taken for this study.

After initial selection, the corresponding slides biopsies were labeled with a fluorescent antibody technique according to the recommendations of the producer. Briefly, green staining was performed with Annexin A3, a marker of low grade GG and normal prostatic tissue, red with p504s, a cytoplasmic marker of prostate cancer.

Annexin A3 and p504s were detected on fixed, paraffin-embedded samples by double immunofluorescence staining. The antigen retrieval was performed by incubating deparaffinized and rehydrated 4-μm thick tissue sections with sodium citrate buffer pH 6.0 in a water bath at 97°C for 30 min. After washing with PBS, non specific binding sites were blocked by incubating the tissue with normal donkey serum (12 μg/mL, 1:5000, Jackson Immunoresearch, West Grove, PA) for 15 min. Following blocking, the sections were washed with PBS and incubated for 1 h at room temperature with monoclonal rabbit anti-P504S (prediluted, clone 13H4, Thermo Fischer Scientific Illkirch, France) and monoclonal mouse anti-annexin A3 (0.47 μg/mL, 1:5000, TgC7 ProVII5C5, ProteoSys AG). The slides were washed in PBS and incubated for 30 min with Alexa Fluor 488 donkey-anti-mouse (10 μg/mL, 1:200, Invitrogen) and Alexa Fluor 568 donkey-anti-rabbit (10 μg/mL, 1:200, Invitrogen). After washing with PBS, the slides were finally mounted with Glycergel (Dako). Substitution of the primary antibodies by the appropriate isotype served as negative controls.

Quantitative digital pathology

The digital imaging system developed by IMSTAR (France), the Pathfinder™ reader analyzer was used integrating the SmartCapture and PathoScan Markindex software modules. All slides were automatically scanned and captured at x10 magnification in green and red fluorescence modalities using specific Alexa 488 (Annexin 3 green) and Alexa 568 (p504 red) filters. The tumor area was automatically detected by its red fluorescence level. In all scored slides, the background was very heterogeneous between specimens. For this reason, the following approach was applied:

i) On the digital, high resolution image of the full specimen, selection by the user of a normal tissue area, of which a threshold was performed to detect automatically the whole tissue section.

ii) Automated detection of the total glandular and interstitial tissue, within which the tumor areas are recognized and displayed with their contour.

iii) Selection by the user of a small area of normal glandular tissue and normal interstitial tissue with high level of Annexin 3.

Consequently, the software measures automatically the level of annexin 3 in normal glandular tissue Fluorescence (NGF) and the background level of fluorescence (BGF).

After this step, based on this information, the software calculates automatically the following parameters: normal tissue area, tumor tissue area, % of tumor/normal area within the sample, annexin 3 mean fluorescence intensity in tumor glandular tissue, and index of intensity in normal glandular tissue corrected for its background, defined as NIG=NGF-BGF, as an internal calibration per specimen.

This index called QGS= Quantitative Gleason Score, was obtained by automated analysis software.

In order to compare the QGS index with the Gleason score, a 0 to 5 scale was used with the following convention:

Normal glandular tissue = QGS 0,

tumour tissue GS 6(3+3)= QGS 1,

tumour tissue GS 7(3+4)= QGS 2,

tumour tissue GS 7(4+3)=QGS 2,

tumour tissue GS 8 (4+4) = QGS 3,

tumour tissue GS 9 (4+5) and 9 (5+4) = 4,

tumour tissue GS 10 (5+5) = 10

Statistical analysis

Kappa statistics were done with medcalc software version 13.2.2 (MedCalc Softwaren Acacialaan 22, B-8400 Ostend, Belgium.

Results

Pathologist analysis

We achieved after a first lecture inter-observer consensus of 96%. In all, except one biopsy, all three pathologists attributed the same GS. In one biopsy existed a discrepancy between GS 6 and 7, after a new lecture, all pathologists attributed the same GS 7 (3+4) to the slide.

Concerning the distribution of GS, no GG 1 or 2 was given on any of the biopsies. Each biopsy was assigned from GS 6 to 9. The distribution of the pathology score is shown on Table 1.

Briefly, 11 biopsies displayed GS 6 (3+3), 7 GS 7, amongst them 4 cases 7 (3+4) and 3 cases 7 (4+3), 6 GS 8 (4+4), no GS 8 (3+5) or 8(5+3) was seen on the samples, and 2 GS 9 (4+5), no GS 9 (5+4) was observed.

Computer analysis

In the computer analysis 8 cases were considered as GS 6, 8 as GS 7, 8 GS 8 and 2 GS 9

The computer analysis was not supposed to make a difference between GS 7 (3+4) and 7(4+3), nor between 9(4+5) and 9 (5+4). Furthermore no case 9 (5+4) was present in our series according to the initial pathology report.

Comparison between pathologists and computer

In 13 (50%) cases, a 100 % agreement existed between pathologists and the computer analysis. In the other 50% a different GS was attributed. In 8 (31%) cases, GS assigned by pathologists was lower than that assigned by the computer. In 5 cases (19%) pathologists attributed a higher GS than the computer. Nevertheless we never detected a difference of more than one GG between human and machine.

When comparing each grade we found in the GS6 group 5 cases (42.3%) of agreement, the other cases were upgraded by the computer, but never more than 1 GG.

In the GS 7 group, the agreement between pathologists and computer analysis was 28.5%, with another 28.5% of cases upgraded and 44% downgraded, once again the difference of the GG was never more than 1 GG (Figure 1).

In the GS 8 cases, an agreement of 83% with 5 cases out of 6 could be achieved (Figure 2). One case was upgraded to GS 9.

In the last group with 2 cases of GS 9 50% consensus was achieved. In 1 case the computer downgraded 1GG the pathologists' consensus.

When comparing Gleason groups according to the previous literature (GS6, 7 and 8-10) the overall agreement between pathologists and computer was 62% (16 cases out of 26). Major discrepancy existed between the GS 6 and 7 (Table 2).

Statistical analysis

Mean inter-observer (pathologists versus computer) weighted kappa for GS groups was 0.532, Standard error 0.106, 95% [0.325-0.739].

Discussion

The application of Gleason grading has changed during the last 40 years, and several major consensus meetings have tried to standardize modern practice of the application of GG. A major meeting was in 2005, organized by the International Society of Urological Pathology, another was held in 2014, and consensus were taken into the WHO classification 2016 of genitourinary tumours [2,4]. Although important efforts were made, a perfect and standardized GG is still not achieved by the pathology community, the problem of treatment and including patients into nomogramms according to the GS still not completely resolved.

Already in earlier times before the ISUP 2005 conference meeting, several inter-observer studies had been made. In 2001, nine confirmed uropathologists evaluated 46 stained slides of prostate needle biopsies. The overall weighted kappa coefficient for GS of the pathologist compared with each remaining pathologist ranged from 0.56- 0.70. The overall kappa coefficient for each pathologist compared with the others for the different GS groups (2-4, 5-6, 7 and 8-10) ranged from 0.47-0.64, which has to be considered as moderate-substantial . At least 70% of the urologic pathologists agreed on the GS group in 38 consensus cases. The pathologists disagreed frequently about low grade carcinomas, tumors with small cribriform proliferations and histological aspects on the border between the grades [7].

The same consensus cases were distributed to 41 pathologists. The overall kappa coefficient was 0.435 with a range from 0.00-0.88, a consistent undergrading of GS 5-6 and 7 with 47% was observed, and especially GG 4 was frequently undergraded (21%) [3]. Since these studies, severalGleason consensus meeting was held, and modified criteria of ill-defined, cribriform glands were redefined as pattern 4 [2]. A shift towards a higher GS was obvious after this conference. Nevertheless the meeting of 2005 was supposed to uniform the interpretation of GG among the pathologists and improve its reproducibility [8]. The latest meeting in 2014 suggested GG groups from I to V, which would permit to have a lowest grade I, which is more obvious also for patients. The problem of the small cribriform proliferation was also resolved during this last consensus meeting, these lesions are considered as GG 4 in the new WHO classification 2016 [4,7].

In our study the consensus not for GG group, but GS between pathologists and computer was observed in 50%, which was a little bit lower than in the above mentioned study, but our results with a mean kappa of 0.532 were better than in the second study. When compared in Gleason groups (6, 7 and 8, 9+10), the agreement was 62%, which joins the findings of the above mentioned studies. The most discordant results were like in the other study between GG3 and 4.

Recently, a group of 15 experts in uropathology analyzed independently a series of needle biopsies with PCa. Of the original 25 cases, 15 (60%) were consensus cases, which meant than a two third majority gave the same GS, no difference was made among GS7 (3+4) and GS7 (4+3). Nevertheless in no case a 100% consensus could be achieved, one case had a 93% agreement. Mean interobserver weighted kappa for GS groups was 0.43. A key problem was to agree on minimal criteria for small foci of GG4 [9].

Another recent study with the same cases as the previous study amongst 337 European pathologists found an agreement between expert and majority members GS in 12 of 15 cases, in three cases members upgraded. Mean GS attributed by the members was higher than in the genitourinary expert group in 9 out of 15 cases. Agreement between consensus and member was 71.4% in GS 6 cases and 56.4% in GS 7 cases. There existed a clear trend to overgrade prostate biopsies by the members [10].

Our results compared to these two studies are better when taking the interobserver kappa, which in the first study was definitively lower (0.43 *vs* 0.532), furthermore the results of GS are slightly better with 62% when comparing consensus cases. We would also like to underline that in no case the GG was more than 1 grade discordant, which joins findings in the literature or is even better than in some studies. Furthermore in the study of Egevad et al. [9] no 100% consensus was achieved amongst all the uropathologists, on the contrary in our study, which is restricted, because it only concerned 4 observers (3 pathologists and 1 computersystem) we had a 100% consensus in 50% of our cases.

GS 8 is considered as an aggressive carcinoma, it has to be recognized as agressive PCa, the agreement of 83% in the GS 8 group was excellent in our study, which means that highly aggressive disease can be recognized and the recognition eventually automatized, which is one of the important findings of this study. Recently Dong et al compared the new GG system to the former and demonstrated that the classical GS 7(4+3) and 8 modified have a development which is close [8]. Both scores should be considered as aggressive disease. In our study an upgrading of the computer system in 11.5% from Gleason 7 to 8 (2 cases) and 8 to 9 (1 case) was observed; Upgrading from GS 8 to 9 does not considerably change the treatment of the patient, as his disease has to be considered as aggressive. According to the findings of Dong, the difference in the evolution of a PCa 7(4+3) and 8(4+4) might not be too discordant [8].

Changing of definition of grades are of course problematic as the prognostic and clinical impact will change over time [11-13]. Our results join the findings of the major consensus studies. Although this is a pilot study and cases are limited, our data clearly show that computer based analysis of GG can be possible. The agreement between the pathologists and the computer system is acceptable with a weighted kappa of 0.532, and a 100% agreement of 50%, which is better than in several inter-observer studies. One bias might have been that the pathologists were used to work together and therefore the interobserver discrepancy was not very high.

There are still discrepancies in the interpretation of the GS among the pathologists. The reasons are multiple, in a recent article, Berney et al. have shown that only 58% of pathologists include a tertiary GG on needle biopsies, only (6% give GS for each score/slide, and 77% give a global GS in the conclusion [11]. Misinterpretation of ISUP is still widely spread; therefore an automatization on slides which have been agreement cases among experts could improve the treatment of patients with PCa.

Considering that the pilot study has already shown a good correlation between GS given by the pathologist and the one found by the Pathfinder system , it would be interesting to include more cases in this study, to refine the computer reading . This will permit to include patients more accurate and consistent groups, and improve the management of patients.

Acknowledgement

To Annette Lesot for technical work.

References

1. Institut de la veille sanitaire. Communiqué de presse - Evolution of the incidence and cancer mortality in France between 1980 and 2012; 2013.
2. Epstein JI, Allsbrook WC Jr, Amin MB, Egevad LL, ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am J Surg Pathol. 2005; 29: 1228-1242.
3. Allsbrook WC, Mangold KA, Johnson MH, Lane RB, Lane CG, Amin MB, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. Hum Pathol. 2001; 32: 74-80.
4. WHO classification of Tumours of the Urinary System and Male Genital Organs. 4th edition ; 2016.
5. Schostak M, Schwall GP, Poznanović S, Groebe K, Müller M, Messinger D, et al. Annexin A3 in urine: a highly specific noninvasive marker for prostate cancer early detection. J Urol. 2009; 181: 343-353.
6. Jiang Z, Woda BA, Rock KL, Xu Y, Savas L, Khan A, et al. P504S: a new molecular marker for the detection of prostate carcinoma. Am J Surg Pathol. 2001; 25: 1397-1404.
7. Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, et al. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. Eur Urol. 2016; 69: 428-435.
8. Dong F, Wang C, Farris AB, Wu S, Lee H, Olumi AF, et al. Impact on the clinical outcome of prostate cancer by the 2005 international society of urological pathology modified Gleason grading system. Am J Surg Pathol. 2012; 36: 838-843.
9. Egevad L, Algaba F, Berney DM, Boccon-Gibod L, Compérat E, Evans AJ, et al. Interactive digital slides with heat maps: a novel method to improve the reproducibility of Gleason grading. Virchows Arch Int J Pathol. août 2011; 459: 175-182.
10. Egevad L, Ahmad AS, Algaba F, Berney DM, Boccon-Gibod L, Compérat E, et al. Standardization of Gleason grading among 337 European pathologists. Histopathology. janv 2013; 62: 247-256.
11. Berney DM, Algaba F, Camparo P, Compérat E, Griffiths D, Kristiansen G, et al. Variation in reporting of cancer extent and benign histology in prostate biopsies among European pathologists. Virchows Arch Int J Pathol. 2014; 464: 583-587.
12. Berg KD, Toft BG, Røder MA, Brasso K, Vainer B, Iversen P. Prostate needle biopsies: interobserver variation and clinical consequences of histopathological re-evaluation. APMIS. 2011; 119: 239-246.
13. Pierorazio PM, Walsh PC, Partin AW, Epstein JI. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. BJU Int. 2013; 111: 753-760.