The Astrophysical Journal, 578:L31–L35, 2002 October 10

BROAD ABSORPTION LINE QUASARS IN THE EARLY DATA RELEASE FROM THE SLOAN DIGITAL SKY SURVEY

Alin Tolea, Julian H. Krolik, and Zlatan Tsvetanov,
Department of Physics and Astronomy, Johns Hopkins University, Charles and 34th Street, Bloomberg Center, Baltimore, MD 21218-2686
Received 2002 May 31; accepted 2002 September 9; published 2002 September 17

ABSTRACT

A new broad absorption line (BAL) quasar sample is derived from the first data released by the Sloan Digital Sky Survey. With 116 objects, it is the largest BAL quasar sample yet assembled. Over the redshift range $1.8 \leq z \leq 3.8$, the crude fraction with broad absorption in the C iv line is $\simeq$15%. This fraction may be subject to small selection-efficiency adjustments. There are also hints of redshift dependence in the BAL quasar fraction. The sample is large enough to permit the first estimate of the distribution of the "balnicity index": subject to certain arbitrary parameters in the definition of this quantity, it is very broad, with (roughly) equal numbers of objects per logarithmic interval of balnicity. BAL quasars are also found to be redder on average than non-BAL quasars. The fraction of radio-loud BAL quasars is (weakly) consistent with the fraction of radio-loud ordinary quasars.

Subject headings: galaxies: active — quasars: absorption lines

1. INTRODUCTION

Broad absorption line (BAL) quasars are one of the most enigmatic varieties of quasars. Resonance lines of ordinary ions—H i, C iv, N v, O vi, Mg ii, and others—are seen in absorption that spreads, often in highly irregular fashion, as much as 60,000 km s^{-1} from line center in the quasar rest frame to the blueward. Previous surveys (e.g., the Large Bright Quasar Survey [LBQS]; Weymann et al. 1991) have shown that BAL quasars, while a minority, are not rare; a population fraction of ~10% is typically estimated. Because few of their other properties are grossly different from ordinary quasars, it is generally thought that all quasars have BAL material, but it covers only a fraction of a solid angle around the quasar nucleus (Weymann et al. 1991). However, subtleties of selection can complicate the inference of the covering fraction from the population fraction (Goodrich 1997; Krolik & Voit 1998).

Numerous technical difficulties have retarded growth in our understanding of BAL quasars. Known cases are relatively rare, numbering less than ~100, not solely because they are a minority of the general population but also because they are readily found only when their characteristic features are redshifted from their rest-frame wavelengths in the ultraviolet into the visible band. Consequently, only those quasars found in somewhat special redshift intervals can be easily searched for broad absorption. It is hard to statistically characterize those BAL quasars that are found because the methods used to discover them often involve some level of subjectivity that is hard to quantify. Even if their selection were easier to articulate, there appears to be so much variation in their properties (profile shapes, relative line strengths, etc.) that it is hard to grasp which properties are generic and which are "accidental."

The quasar sample being compiled by the Sloan Digital Sky Survey (SDSS; York et al. 2000) offers a way out of this impasse. When complete, it will be both very large (~10^5 in all) and selected in a uniform and quantifiable manner. In future work, we hope to present a statistical analysis of BAL quasars in this entire sample. Here we offer a preliminary installment on this project in the form of a more modest BAL quasar sample drawn from the first data released from the project to public view, the early data release (EDR; Stoughton et al. 2002).

Several collections of BAL quasars have already been drawn from early SDSS data (Menou et al. 2001; Hall et al. 2002); these were, however, oriented toward the "by-eye" selection of small subsamples that were special in some way (radio-loud in the former case, extraordinary profiles in the latter). The work reported here differs in that it is the first attempt to create a systematically selected sample from the SDSS.

From the EDR, Schneider et al. (2002) created a quasar catalog containing 3814 quasars, selected (mostly) on the basis of their location in four-color space and on a (mostly) uniform *i*-magnitude limit. In order to present more clearly defined statistics, we have refined this sample so that it is almost homogeneously selected (see § 2.1). Within that subsample (about 80% of the full EDR quasar catalog), roughly one-quarter (796) fall within the redshift range within which it is feasible to search for C iv BAL features.

With an eye toward the homogeneity of selection to be achieved in the full SDSS, we invented an automated BAL selection algorithm that processes SDSS spectral data in a uniform way and identifies BAL quasars in a uniform manner (see § 2.2). Using this algorithm, we have identified 116 BAL quasars, whose statistical properties are discussed in § 3. Although the EDR represents a tiny fraction of the ultimate SDSS quasar sample, the BAL sample so derived is now the largest (as well as the most homogeneously selected) such sample known.

2. DETAILS OF SAMPLE SELECTION

2.1. *Quasar Selection*

The EDR quasar catalog (Schneider et al. 2002) was compiled by applying several different selection criteria (see Stoughton et al. 2002 for details). Most of its objects were chosen on the basis of colors lying outside the "stellar locus" in the four-color space formed by the SDSS five-filter (*u*, *g*, *r*, *i*, and *z*) photometry.[1] However, roughly 20% of the quasars in this list were chosen in a much less well-defined fashion (in the jargon of the SDSS, these were selected for the "serendipity" sample). In addition, much smaller numbers of quasars

[1] Technically, because the photometry published in the EDR had not received final calibration, the magnitudes were shown as i^*, etc. In this Letter, all such magnitudes should be understood in that sense, although we will forgo making the distinction explicit.

were selected not on the basis of their photometric colors but because they were close in the sky to known radio sources in the Faint Images of the Radio Sky at Twenty cm (FIRST; Becker, White, Helfand 1995) catalog or X-ray sources in the *ROSAT* All-Sky Survey (Voges et al. 1999). For the purposes of this Letter, which concentrates on statistics, we have pruned the EDR quasar catalog to include *only* those flagged as quasar candidates by a color-based targeting algorithm.

Because the EDR data were compiled during the test year of the project, the color-based selection algorithm was not the same throughout. With regard to issues relevant here, the variations can be reduced to two slightly different versions, very nearly equal in sky coverage. In both, the primary flux limit was $i \leq 19$ mag.[2] Both versions also shared the same primary color criterion: select only those objects whose colors lie at least 4 σ away from the stellar locus. An exception was made to this rule in order to cope with the fact that at redshifts $z \simeq 2.5$–3.0, quasars have colors nearly indistinguishable from those of A- and/or F-type stars (Fan 1999; Richards et al. 2001). In this portion of the stellar locus, quasar candidates were selected, but at lower efficiency (Stoughton et al. 2002; G. T. Richards 2002, private communication).

The two versions differed in two ways: One version rejected objects with colors approximating those of A-type stars; in the other, objects with colors similar to those of hot white dwarfs or unresolved M dwarf–white dwarf pairs were also removed from the quasar candidate list. In addition, in those runs in which white dwarf–like colors were rejected, a special procedure was adopted in order to enhance sensitivity to quasars with $z \geq 3.5$. To find more of these quasars, the magnitude limit was relaxed to $i = 20$ mag in the region of four-color space where previously located high-redshift quasars were found. We will show in § 3.1 that these variations in quasar candidate selection had negligible effects on the discovery of BAL quasars.

Quasar candidates were labeled as "quasars" by the spectroscopic pipeline if the cross-correlation between their spectra and a quasar template spectrum was greater than the cross-correlation with any of the other templates (stars, galaxies, etc.). For confirmation, objects were required to pass two further tests: that their spectra possess at least one emission line with FWHM ≥ 1000 km s^{-1} and that their absolute magnitude $M_i \leq -23$ (for $H_0 = 50$ km s^{-1} and $q_0 = 0.5$).

In the full EDR quasar catalog, there were 3814 objects. Our sample contains only the 3107 identified by one of the two versions of the color-selection rules.

2.2. *BAL Identification*

The C IV line is centered at 1550 Å in the rest frame, so it appears in the SDSS spectra (which nominally cover the wavelength range from 3900 to 9100 Å) only for redshifts $1.5 < z < 4.9$. However, several effects limit this range further. First, although the nominal blue cutoff is 3900 Å, in practice, throughput, and therefore the signal-to-noise ratio (S/N), drop sharply shortward of ≃4100 Å and (more gradually) longward of ≃8000 Å. Moreover, in order to measure possible absorption to the blue of line center, we must be able to see some line-free continuum redward of the emission line center and also follow the line far enough to the blue that we are confident we have defined the entire absorption profile. These requirements restrict the permissible range of redshifts to roughly $1.8 \leq z \leq 3.8$, cutting our sample size to 796 objects.

Redshifts supplied by the SDSS spectroscopic pipeline are typically accurate to ~1000 km s^{-1}, which suffices for the cut described in the previous paragraph but is not accurate enough for absorption-line measurements. These measurements require greater accuracy because the classical definition of BAL quasars (Weymann et al. 1991) counts only absorption at least 3000 km s^{-1} to the blue of line center in the rest frame.

The C IV emission line cannot be used to define the redshift to this level of accuracy because the very absorption we are interested in can cut into the emission line so severely that it is unclear where the line center occurs. Instead, we define the quasar rest frame in terms of the emission line C III] λ1909.[3] To determine the redshift this way, we fitted a Gaussian plus a linear component to the measured flux in the (pipeline redshift) rest-frame wavelength range of 1860–1960 Å. We take the center of the Gaussian in the best fit to define the true observed wavelength for rest frame of 1909 Å.

To search for absorption—which can be extremely difficult in these objects, in which emission and absorption features can occupy most of the spectrum—one must first locate the continuum. Our solution to this problem is to first fit a power law to the continuum data in five line-free windows: 1790–1820, 1975–2000, 2140–2155, 2240–2255, and 2265–2695 Å. Holding that component fixed, we then fitted a half-Gaussian to the red half of the C IV emission feature lying above the fitted continuum, taking the line center as 1549.5 Å in the rest frame (this wavelength was derived by equally weighting the two components of this doublet). We deliberately ignore the blue half of the emission line so as to avoid confusion by absorption; the continuum windows are chosen so as to avoid contamination by the He II λ1640, C III] λ1909, and Mg II λ2800 emission lines as well as from various Fe II emission complexes. We then extrapolate the power-law portion of this fit to define the continuum blueward of 1550 Å.

The final step in our procedure is to compute the "balnicity index" (BI) for each quasar in this redshift range, following the definition given in Weymann et al. (1991):

$$\mathrm{BI} = \int_{-25{,}000}^{-3000} dv \left(1 - \frac{F_\lambda}{0.9C_\lambda}\right) C, \tag{1}$$

where the measured flux per unit wavelength is F_λ, the extrapolated fitted continuum is C_λ, and C is a function whose value is unity when the quantity between parentheses has been positive for at least 2000 km s^{-1} to the red of the current wavelength and zero otherwise. The lower limit on the integral is designed to avoid confusion with the Si IV line; the upper limit on the integral is designed to exclude associated absorption. A comparison is made with 0.9 times the continuum rather than the full continuum in order to be conservative with respect to noise features. The function C ensures that only truly broad features are counted. In effect, the BI amounts to a sort of equivalent width. Following Weymann et al. (1991), we declare a quasar to be a BAL when its BI is greater than zero.

[2] Objects were also required to be fainter than a limit chosen to avoid image saturation and cross talk between spectroscopic fibers. This was $i = 15.0$ mag in some cases, $i = 16.5$ mag in others; because so few objects are near the bright limit, the change makes essentially no difference to sample statistics.

[3] According to Vanden Berk et al. (2001), this line is, on average, offset only ≃200 km s^{-1} from the systemic host redshift as defined by [O III] λ5007. For similar reasons, Weymann et al. (1991) used a weighted mean of the C IV, C III, and Mg II emission lines to determine their redshifts.

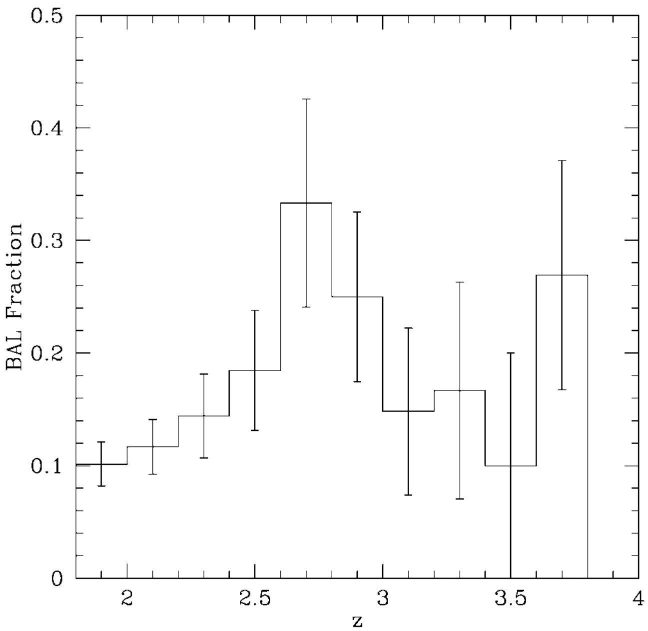

FIG. 1.—BAL fraction in the sample as a function of redshift. Error bars are 1 σ and purely statistical.

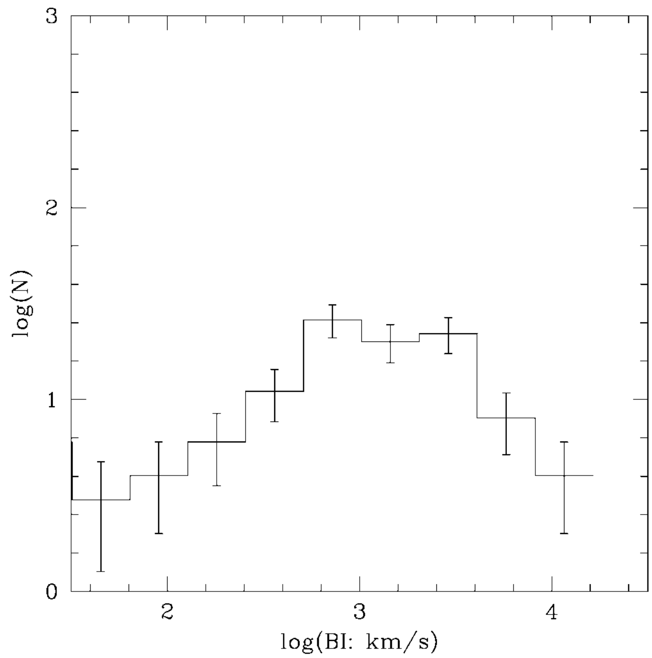

FIG. 2.—Distribution of the BI in the sample (*solid histogram*). The binning is logarithmic in the BI. Error bars are 1 σ.

3. RESULTS

In view of the preliminary nature of this sample, here we present only a few tentative results. The statistical and systematic uncertainties in the numbers presented here will be substantially reduced in the far larger full SDSS sample; that sample will also enable other, more detailed studies.

3.1. *BAL Fraction*

Of the 796 quasars in the redshift range where we can search for BAL quasars, we find that 116, or $\simeq$15%, are C IV BAL quasars. The apparent BAL fraction in this sample is also strongly dependent on redshift (Fig. 1): it is only $\simeq$10% in the redshift range $1.8 \le z \le 2.2$ but rises to $\simeq$33% near $z \simeq 2.7$ and averages $\simeq$20% for $2.2 \le z \le 3.8$.

Some of this redshift dependence may be real, but there are also redshift-dependent systematic effects. Near $z \simeq 1.8$, some BAL quasars may be lost because of the relatively poor S/N at the blue end of the spectrograph. In the range $2.5 \le z \le 3$, distinguishing quasars and stars by color becomes difficult. Both the broad absorption itself and intrinsic differences in continuum shape (§ 3.3) can give BAL quasars colors that are different from ordinary quasars; their selection efficiencies can therefore differ significantly in this redshift range. The spike in the "raw" BAL fraction near $z \simeq 2.7$ may be the result of this differential selection efficiency (see T. Reichard et al. 2002, in preparation, for further exploration of this issue).

On the other hand, dividing the sample according to the color-selection procedure used, we find negligible differences. Both BAL and ordinary quasar selection efficiencies were equal to well within Poisson errors.

The LBQS found a significantly smaller raw fraction (9%), but Weymann et al. (1991) corrected this figure to $\simeq$12% because the BAL itself removed enough flux that many BAL quasars dropped below the survey flux limit. In the SDSS, by contrast, the flux limit is applied in the *i* band, near 8500 Å. Only for redshifts $\simeq$4 would a C IV λ1550 BAL quasar influence the *i*-band flux, and we do not even consider such high-redshift quasars in the sample at hand. Consequently, the SDSS BAL fraction needs to be corrected for this effect only at very high redshift or for the special case of "LoBALs," BAL quasars with absorption in the Mg II λ2800 line. When Mg II absorption is present, the BAL would remove flux from the *i* band when $z \simeq 2.5$.

Overall, then, particularly for $z \ge 2.2$, the SDSS appears to find a somewhat larger BAL fraction than the LBQS. Allowing for the various systematic errors, our best preliminary estimate is a BAL fraction of $\simeq$15%–20% for this redshift range, but possibly nearer to 10%–12% for $1.8 \le z \le 2.2$. We expect these numbers to be refined as the statistics improve and as the systematic effects become better understood. Using the full SDSS quasar sample, it should become possible to search for genuine redshift (or luminosity) dependence in the BAL fraction.

3.2. *Balnicity Index Distribution*

The BI distribution for our sample is shown in Figure 2. The plot shows $\log\,(dN/d\log \mathrm{BI}) = \log\,[\mathrm{BI}\; dN/d(\mathrm{BI})]$; within the loose constraints placed by the relatively small sample size, there are roughly equal numbers of objects in equal logarithmic bins. However, we stress that the shape of this distribution below ~1000 km s^{-1} is strongly dependent on the arbitrary velocity offset parameter used in the definition of balnicity to distinguish "associated absorbers" from truly "broad" absorption (see the discussion in Hall et al. 2002).

The shape of this distribution has several interesting implications. First, because BAL quasars of very small BI are common, the arbitrariness of the velocity offset parameter means that the distinction between weak BALs and associated absorption lines is difficult to mark and that the nominal BI for these objects likely underestimates the "physical" absorption. Second, if we take the definition of balnicity at face value, the breadth of its distribution is consistent with the anecdotal sense of the diversity of BAL profiles derived from previous, smaller

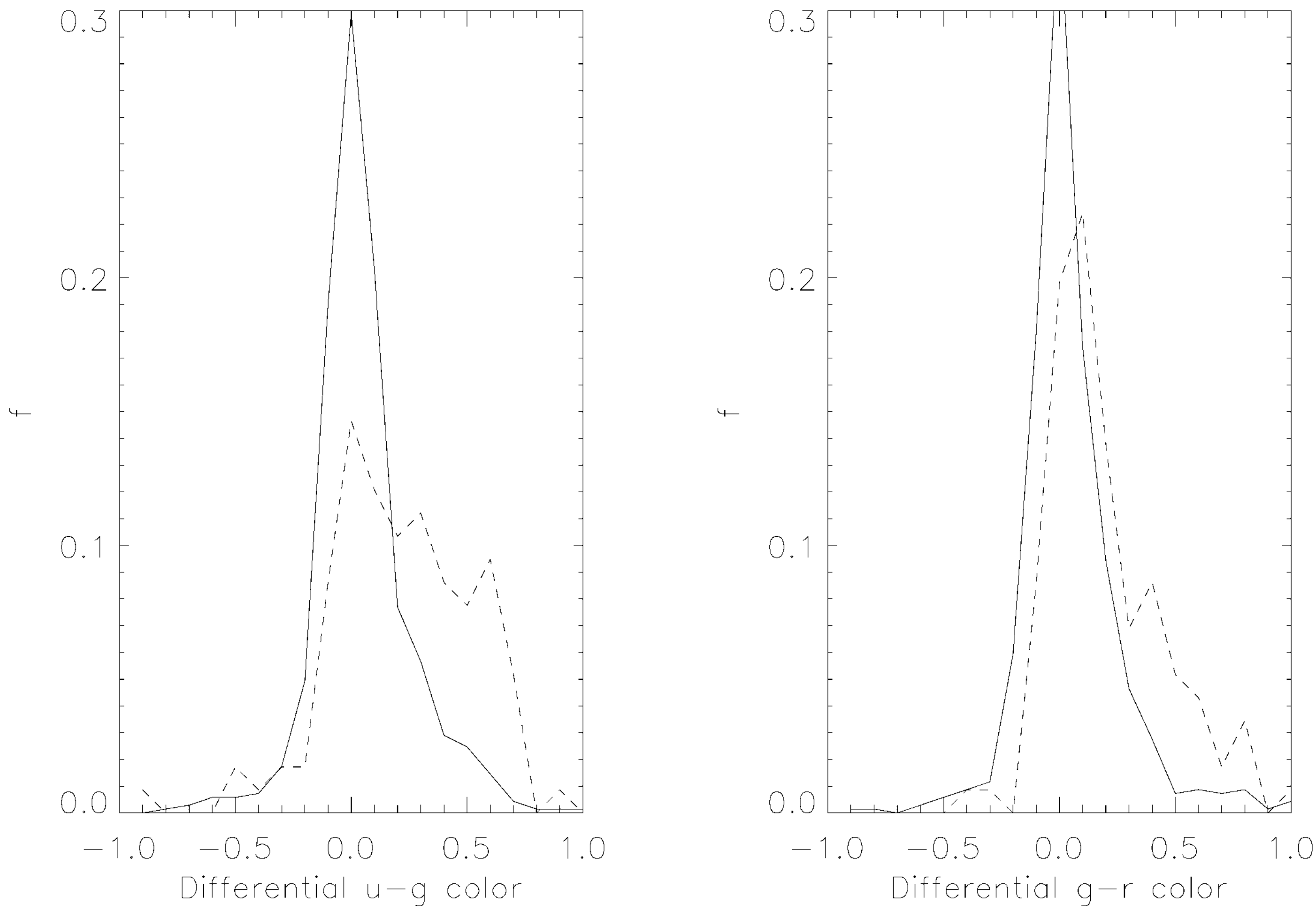

FIG. 3.—The distribution of colors after subtracting off the mean color of quasars at the individual quasars' redshifts. The solid line represents non-BAL quasars; the dashed line represents BAL quasars. The left-hand panel is the distribution for normalized $u-g$; the right-hand panel is the distribution for normalized $g-r$.

samples. Third, the shallow slope of the distribution at the high-balnicity end suggests that the maximum-velocity width of BAL quasars is probably as yet ill-defined.

3.3. *Colors of BAL Quasars*

The mean quasar color is a strong function of redshift (Fan 1999; Richards et al. 2001). To contrast the colors of BAL quasars and non-BAL quasars most clearly, in Figure 3 we show the distribution of colors after subtracting the mean color for our sample at each quasar's redshift. Particularly in $u-g$, the distribution of BAL quasar colors is shifted distinctly to the red (see also Menou et al. 2001). In the mean, the $g-r$ color difference is about 0.18 mag; in $u-g$, it is about 0.34 mag. Both color offsets are crudely constant with redshift for $1.8 \leq z < 3$. This trend is in the same sense (although somewhat smaller than) the color contrast in the FIRST survey (Becker et al. 2000; Brotherton et al. 2001), in which radio-selected BAL quasars were, on average, $\simeq$0.5 mag redder (in a color roughly equivalent to $B-R$) than their radio-selected non-BAL quasars.

There are several possible explanations for this effect. It may be, for example, that there is dust associated with the absorbing material itself. It is also possible that the redder colors are due to dust, but farther from the nucleus along our line of sight (e.g., as proposed by Goodrich 1997).

On the other hand, BAL quasars may differ from ordinary quasars in some intrinsic fashion, perhaps having a different mean ratio of luminosity to Eddington luminosity. Alternatively, we may view them from a special angle. If the optical continuum is generated in an accretion disk, one might expect both that the continuum shape would depend on $L/L_{\rm Edd}$ and that there is wavelength-dependent limb darkening (e.g., Hubeny et al. 2000). The latter effect would create a systematic color offset if the absorbing matter lies in a special direction relative to the disk.

That BAL quasars are preferentially redder than ordinary quasars affects our ability to find them. The reason our BAL fraction is larger than the fraction in the LBQS may be that quasars in the LBQS were selected in part on the basis of blue colors. If so, the comparative lack of color bias in the SDSS may be critical for obtaining a fair estimate of the size and character of the BAL quasar population. In addition, if the redder colors are also associated with continuum flux that is weaker in the BAL quasar direction, the fraction of the sky around the nucleus covered by BAL material would be larger than the population fraction of BAL quasars (Goodrich 1997; Krolik & Voit 1998).

3.4. *Radio-loud Fraction*

We close with a brief comment about the radio properties of these BAL quasars. Weymann et al. (1991) found that none of their BAL quasars were radio-loud and therefore suggested an anticorrelation between the two properties. On the other hand, Becker et al. (2000) argued that, if anything, BAL quasars were *more* likely to be found in radio-selected than in optically selected samples.

Radio loudness is often defined as $R \equiv F_\nu(5\ \mathrm{GHz})/F_\nu(4400\ \mathrm{Å}) \gtrsim 10$–30. This criterion can be applied to only about 4% of the SDSS quasars because radio data are available for only those brighter at 1.4 GHz than the FIRST flux limit. Given the optical flux limit of the SDSS quasar sample, essentially all those quasars in the EDR detected by FIRST are radio-loud by this criterion.[4] Five of the 116 BAL quasars found in our sample are radio-loud by this definition. Our results are therefore consistent with the proposition that there is no difference between the radio-loud BAL fraction and the radio-loud fraction among ordinary quasars. However, in view of the very small number of objects and the incompleteness of radio data for our sample, this conclusion must be tentative at best.

[4] Comparing this radio-loud fraction with the ≃15% found by Kellermann et al. (1989) in the Palomar-Green sample suggests that many radio-loud quasars in the SDSS are a little too faint to have been detected by FIRST. Ivezić et al. (2002) estimate a radio-loud fraction of ≃8%; relative to this fraction, there are still numerous radio-loud quasars in this sample that must fall just below the FIRST detection limit.

We thank Tim Heckman, Gordon Richards, and Tim Reichard for numerous helpful conversations and suggestions. J. H. K. was partially supported by NASA grant NAG5-9187.

REFERENCES

Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559
Becker, R. H., White, R. L., Gregg, M. D., Brotherton, M. S., Laurent-Muehleisen, S. A., & Arav, N. 2000, ApJ, 538, 72
Brotherton, M. S., Tran, H. D., Becker, R. H., Gregg, M. D., Laurent-Muehleisen, S. A., & White, R. L. 2001, ApJ, 546, 775
Fan, X.-H. 1999, AJ, 117, 2528
Goodrich, R. W. 1997, ApJ, 474, 606
Hall, P. B., et al. 2002, ApJS, 141, 267
Hubeny, I., Agol, E., Blaes, O., & Krolik, J. H. 2000, ApJ, 533, 710
Ivezić, Z., et al. 2002, AJ, submitted (astro-ph/0202408)
Kellermann, K. I., Sramek, R., Schmidt, M., Shaffer, D. B., & Green R. 1989, AJ, 98, 1195
Krolik, J. H., & Voit, G. M. 1998, ApJ, 497, L5
Menou, K., et al. 2001, ApJ, 561, 645
Richards, G. T., et al. 2001, AJ, 121, 2308
Schneider, D. P., et al. 2002, AJ, 123, 567
Stoughton, C., et al. 2002, AJ, 123, 485
Weymann, R. J., Morris, S. L., Foltz, C. B., & Hewett, P. C. 1991, ApJ, 373, 23
Vanden Berk, D. E., et al. 2001, AJ, 122, 549
Voges, W., et al. 1999, A&A, 349, 389
York, D. G., et al. 2000, AJ, 120, 1579