



Published in final edited form as:

*J Nerv Ment Dis.* 2013 October ; 201(10): 860–871. doi:10.1097/NMD.0b013e3182a5c184.

## GAP-REACH:

### A Checklist to Assess Comprehensive Reporting of Race, Ethnicity, and Culture in Psychiatric Publications

Roberto Lewis-Fernández, MD<sup>\*,†</sup>, Greer A. Raggio, MPH<sup>‡</sup>, Magdaliz Gorritz, MPH<sup>†</sup>, Naihua Duan, PhD<sup>\*,†,§</sup>, Sue Marcus, PhD<sup>\*,†,§</sup>, Leopoldo J. Cabassa, PhD<sup>\*,†,||</sup>, Jennifer Humensky, PhD<sup>\*,†</sup>, Anne E. Becker, MD, PhD, ScM<sup>¶</sup>, Renato D. Alarcón, MD, MPH<sup>#</sup>, María A. Oquendo, MD<sup>\*,†</sup>, Helena Hansen, MD, PhD<sup>\*\*</sup>, Robert C. Like, MD, MS<sup>††</sup>, Mitchell Weiss, MD, PhD<sup>††</sup>, Prakash N. Desai, MD<sup>§§</sup>, Frederick M. Jacobsen, MD, MPH<sup>|||</sup>, Edward F. Foulks, MD, PhD<sup>¶¶</sup>, Annelle Primm, MD, MPH<sup>###</sup>, Francis Lu, MD<sup>\*\*\*</sup>, Alex Kopelowicz, MD<sup>†††</sup>, Ladson Hinton, MD<sup>\*\*\*</sup>, and Devon E. Hinton, MD, PhD<sup>†††</sup>

\*Department of Psychiatry, College of Physicians & Surgeons, Columbia University, New York, NY †New York State Psychiatric Institute, New York, NY ‡Department of Psychology, Drexel University, Philadelphia, PA §Department of Biostatistics, Mailman School of Public Health, Columbia University ||School of Social Work, Columbia University ¶Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA #Department of Psychiatry and Psychology, Mayo Clinic College of Medicine, Rochester, MN \*\*Departments of Psychiatry and Anthropology, New York University, New York, NY ††Department of Family Medicine and Community Health, UMDNJ–Robert Wood Johnson Medical School, New Brunswick, NJ †††Department of Epidemiology and Public Health, Swiss Tropical Institute, and University of Basel, Switzerland §§Department of Psychiatry, University of Illinois at Chicago |||Department of Psychiatry and Behavioral Sciences, George Washington University, Washington, DC ¶¶Department of Psychiatry, Tulane University, New Orleans, LA ###American Psychiatric Association, Arlington, VA; Department of Psychiatry, Johns Hopkins Medical School, Baltimore, MD \*\*\*Department of Psychiatry, University of California, Davis †††Department of Psychiatry, University of California, Los Angeles †††Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School

## Abstract

Growing awareness of health and health care disparities highlights the importance of including information about race, ethnicity, and culture (REC) in health research. Reporting of REC factors in research publications, however, is notoriously imprecise and unsystematic. This article describes the development of a checklist to assess the comprehensiveness and the applicability of

Copyright © 2013 by Lippincott Williams & Wilkins

Send reprint requests to Roberto Lewis-Fernández, MD, New York State Psychiatric Institute, Unit 69, 1051 Riverside Dr, New York, NY 10032. rlewis@nyspi.columbia.edu.

The authors are members or invitees of the Cultural Committee of the Group for the Advancement of Psychiatry (GAP), which is an organization of nationally respected psychiatrists and other physicians dedicated to shaping psychiatric thinking, public programs, and clinical practice in mental health.

## DISCLOSURES

The authors declare no conflicts of interest.

REC factor reporting in psychiatric research publications. The 16-item GAP-REACH<sup>®</sup> checklist was developed through a rigorous process of expert consensus, empirical content analysis in a sample of publications ( $N = 1205$ ), and interrater reliability (IRR) assessment ( $N = 30$ ). The items assess each section in the conventional structure of a health research article. Data from the assessment may be considered on an item-by-item basis or as a total score ranging from 0% to 100%. The final checklist has excellent IRR ( $\kappa = 0.91$ ). The GAP-REACH may be used by multiple research stakeholders to assess the scope of REC reporting in a research article.

## Keywords

Race; ethnicity; culture; checklist; psychiatric literature

---

Practitioners in diverse areas of health-related activity—policy, advocacy, research, and clinical practice—have increasingly acknowledged the role that factors related to race, ethnicity, and culture (REC) play in population health and health care delivery (Centers for Disease Control and Prevention [CDC], 1993; International Committee of Medical Journal Editors, 2010; Mir et al., 2012). This growing awareness is caused in part by major demographic shifts in the United States, concerns about the cross-cultural validity of research findings, and discovery of disparities in clinical assessment and health care access and quality across REC groups. Information based on comprehensive and reliable use of REC variables in health research is particularly needed to clarify and redress these disparities (Drevdahl et al., 2006; Institute of Medicine [IOM], 2009; Kahn, 2003; National Partnership for Action to End Health Disparities, 2011).

Reporting of REC factors in health research publications, however, is notoriously imprecise and unsystematic (Bhopal and Kohli, 1997; Mak et al., 2007; Sankar et al., 2007). Problems include lack of consistency in defining, operationalizing, and differentiating the constructs “race,” “ethnicity,” and “culture”; lack of explanation about how respondents are classified into REC categories; poor or absent justification for including or omitting REC information; and low utilization of REC factors in data analyses and interpretation of results (CDC, 1993; Comstock et al., 2004; Drevdahl et al., 2006; Mak et al., 2007; Sankar et al., 2007; Williams, 1994). These deficiencies suggest a persistent lack of consensus about how to include REC factors in health research reporting and arguably in the general design and interpretation of health research.

For nearly 2 decades, US federal agencies, journal editors, organizations and leaders in cultural mental health, and other groups have attempted to rectify the situation. Since 1993, the NIH and the Department of Health and Human Services (DHHS) have issued guidelines for investigators mandating the use of standardized categories to obtain information on race/ethnicity from research participants (DHHS, 1997; NIH, 1993, 1994, 2001). Also in 1993, the CDC called for improvements in the justification, measurement, interpretation, and limitations of the collection and reporting of race/ethnicity data (CDC, 1993). In 2001, the NIMH launched a 5-year initiative intended to increase the gathering and reporting of research data on REC subgroups with sufficient statistical power to inform clinical practice and mental health policies (NIMH, 2001). As recently as 2009, the IOM prepared a list of

standardized REC variables for use in all US health reporting (IOM, 2009). In June 2011, the US DHHS announced a public review process of standards based on this IOM report before their implementation in the Patient Protection and Affordable Care Act for national health care reform (DHHS, 2011). Likewise, US and international journal editors have repeatedly requested greater justification for and standardization in the use of REC variables (American Academy of Pediatrics Committee on Pediatric Research, 2000; International Committee of Medical Journal Editors, 2003, 2010; Rivara and Finberg, 2001; Salway et al., 2011a).

However, the impact of these efforts on US psychiatric research remains unclear. Surveys on the reporting of REC variables in mental health research have shown only limited improvement since 1990. Between 1990 and 1999, the proportion of more than 6000 articles published in six American Psychological Association journals that focused on racial/ethnic minorities or reported REC-centered analyses rose from 3.2% to only 5.3% (Imada and Schiavo, 2005). The number of articles published in *Psychiatric Services (PS)* that “covered issues related to racial and ethnic groups” rose eightfold between 1950 and 1989 but decreased by 23% from its 1989 crest during the following decade (Bell and Williamson, 2002, p. 420). Between 1995 and 2004, more articles reporting on NIMH-funded clinical trials provided some REC information, but the proportion of trials that reported detailed REC data did not increase. Overall, only 48% of the articles provided a complete REC profile of study participants and only 30% included subgroup analyses by REC categories (Mak et al., 2007). To our knowledge, no research on the reporting of REC factors has been published in the US psychiatric literature since 2004.

Limited progress in REC factor reporting may be caused in part by the absence of a systematic and reliable method to assess the scope of how REC factors have been applied in any given psychiatric publication. Although policy makers, investigators, and editorial committees have previously described some of the conceptual domains that characterize comprehensive reporting, very little work has been done to develop a systematic, operational method to account for these domains. Most US-based research has focused on the use and definition of individual concepts (e.g., ethnicity). More systematic, recent work in the United Kingdom by Salway et al. (2011a) has not been widely adopted. Nevertheless, substantial agreement exists about the most important potential domains to consider, such as clear definition of REC variables, rationale for their inclusion in the study, appropriate methods for their ascertainment, and their inclusion in data analysis and interpretation (CDC, 1993; Comstock et al., 2004; IOM, 2009; Kaplan and Bennett, 2003; Williams, 1994; Salway et al., 2011b).

A way forward is suggested by work in other areas of health research, where checklists and scales have been developed to systematize standards for research reporting, including clinical trials, meta-analyses, and literature reviews (Begg et al., 1996; Jadad et al., 1996; Kocsis et al., 2010; Moher et al., 1999, 2009, 2011; Oxman et al., 1991; Stroup et al., 2000). Some of these measures remain descriptive, whereas others, such as the Consolidated Standards for Reporting of Trials (CONSORT) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRESMA) statements, provide standards for research reporting (Begg et al., 1996; Moher et al., 2009). Adoption of these guidelines by

editors has resulted in improved standardization and quality of research reports (Al Faleh and Al-Omran, 2009; Moher et al., 2001; Plint et al., 2006).

The Cultural Committee of the Group for the Advancement of Psychiatry (GAP) undertook the development of a checklist operationalizing the criteria for the assessment of comprehensive reporting of REC factors in psychiatric publications. This article describes the development of the checklist, known as GAP-REACH<sup>®</sup> (Race, Ethnicity, And Culture in Health), based on expert consensus, content analysis of a psychiatric publication sample, and item-based reliability assessment. To develop and test the GAP-REACH, we sampled articles published during the period of January 2000 to December 2002; presumably, most of these were drafted and submitted before the 2001 NIMH initiative that focused on increasing REC factor reporting. This allows us to establish a baseline of REC factor reporting in the early 2000s, against which we can, in the future, assess changes over time in a content-analysis study of publications subsequent to the 2001 NIMH initiative.

For each of the main components of a research article, the checklist asks the following questions: Were REC factors included? If so, were the minimum suggested research standards achieved (*e.g.*, terms defined, methods described)? If REC factors were not included, was it because these were inapplicable to the study methodology? It is important to note at the outset that the GAP-REACH does not assess whether inclusion of REC factors is relevant to the topic of the article. This would require a more complex assessment that would vary across research topics and depend heavily on past findings (*i.e.*, previous studies may have shown that REC factors are not relevant to the topic). In this sense, the checklist strives to be applicable, although possibly not always relevant, to all types of psychiatric publications. Subsequent investigations will explore whether the checklist can evolve into a set of guidelines for REC reporting that also take into account the relevance of REC factors to the research topic.

## METHODS

The process of developing the checklist followed four stages, illustrated in Figure 1.

### Stage 1: Development of List of Domains and Data Abstraction Form

Key domains that characterize a comprehensive examination of REC factors in psychiatric research publications were proposed by members and invitees of the GAP Cultural Committee, via expert consensus, until no new domains emerged. The following criteria guided domain selection: conceptual relevance, applicability across psychiatric subareas, and consistency with recommendations from previous surveys of REC factors. Because we expected articles to vary substantially in use and analytic treatment of REC factors, we included a combination of domains that ranged from very simple (*e.g.*, article includes any REC-related word, such as “ethnicity”) to more complex (*e.g.*, article reports psychometric data for each REC subgroup in the sample).

The list of domains was organized under four categories corresponding to conventional sections of research reports (Introduction/ Background, Methods, Results, and Discussion), plus an overall category. This format is consistent with the content organization of other

checklists used to assess research reporting (Moher et al., 1999; Stroup et al., 2000). The final domains were formulated as questions on a data abstraction form for REC-related information. The purpose of the abstraction was to assess whether the proposed domains were feasible to ascertain and apply across articles with diverse topics and methods. In subsequent stages, the abstracted data were used to refine the conceptual domains into specific items and response options in the final GAP-REACH checklist. Detailed instructions aimed at standardizing data abstraction were drafted for each question.

## Stage 2: Data Abstraction

In this stage, we applied the abstraction form to a sample of psychiatric research articles. We oversampled articles that were more comprehensive in their use of REC factors to maximize the yield of substantive answers used to refine the checklist domains into final operational criteria. Checklist domain 1 (the inclusion of REC terms in the title or abstract) was used as a screener to select potentially “REC-focused” articles for in-depth abstraction. We reasoned that articles that mentioned REC factors in the title and/or abstract would be more likely to emphasize these factors in the text (Barley and Salway, 2008). A cohort of articles that were not REC-focused was also selected for in-depth abstraction. This second cohort allowed us to test whether our screening procedure successfully identified articles focused on REC factors, by estimating the correlation between domain 1 and domains 2 to 16.

Articles were considered REC-focused when a) the title or abstract contained any of the following words: *race*, *ethnicity*, *culture*, or any related term(s) (e.g., *racial*) or b) the title or abstract referred to one or more racial/ethnic categories (i.e., black/African-American) or ethnonational groups (i.e., Japanese-Americans). Articles that lacked any REC terminology in the title or abstract were classified as “non-REC focused.”

## Journal Selection

We selected seven psychiatric journals to generate our publication sample for data abstraction: *American Journal of Psychiatry (AJP)*, *Archives of General Psychiatry (AGP)*, *Biological Psychiatry (BP)*, *Journal of the American Academy of Child and Adolescent Psychiatry (JAACAP)*, *Journal of Clinical Psychiatry (JCP)*, *Journal of Nervous and Mental Disease (JNMD)*, and *PS*. These journals represent a variety of research areas in psychiatry and are all well-recognized publications. The goal was to assemble a purposive sample comprising journals both representative of mainstream psychiatry and also diverse in research content focus (e.g., biological psychiatry, services research). Each journal is published at least monthly, which ensured an adequate sample size of articles within a specified period for checklist development.

## Selection of Articles

Articles were included if published in any of the seven selected journals between January 2000 and December 2002 as original research reports with at least 10 participants. (No special issues dedicated to REC factors in any of the journals surveyed were found during this time frame.) Animal studies, reviews, case reports, editorials, commentaries, and letters were excluded.

To enhance cross-journal applicability of the checklist, we stratified the REC-focused articles by journal. For each journal, we selected up to 15 REC-focused articles using our criterion for potentially REC-focused articles described above, starting from the December issue of 2002 and working backward until the quota of 15 REC-focused articles per journal was reached. For journals that did not reach the quota, the process stopped when all articles from the 3-year period were screened.

Once article screening was completed, to serve as the comparison group, a random sample of non-REC-focused articles from each journal was selected during the journal-specific screening time frame that yielded the REC-focused sample (*e.g.*, 6 months). Each journal contributed an equal number of REC-focused and non-REC-focused articles for the analysis; this strategy served to maximize the precision of the correlation estimate between domains 1 and 2 to 16.

Screening of articles was performed by four masters-level research assistants. Interrater reliability (IRR) of coding results for this domain was assessed with a 20% random sample of articles. Reviewers were blinded to the initial coding results.

### **Abstraction**

Using the abstraction form developed in stage 1, two GAP committee members independently extracted the data from the REC-focused and non-REC-focused articles for domains 2 to 16. The instructions were refined iteratively during in-person meetings, e-mail interactions, and conference calls, on the basis of group discussion of articles that presented ambiguities. The first author then compared the two sets of results to identify and resolve discrepancies and ensure uniform classification (Williams, 1994). Remaining discrepancies were resolved through committee consensus. The committee members were instructed to identify domains that were not feasible to implement during the abstraction process.

### **Stage 3: Development of the Checklist**

Once all articles were abstracted, the GAP committee used the results to compose a checklist with individual items representing each domain and to develop scoring instructions for the final checklist. The goals at this stage were parsimony, clarity, and ease of use of the resulting items. Items deemed inapplicable to a particular article were noted.

This process resulted in the final 16-item version of the checklist. The percentage of inapplicability of each item was calculated for each journal and compared across the journals using separate chi-square tests. A scoring scheme was devised to provide a total score that accounted for the possibility of item inapplicability. To assess the internal consistency of the checklist items, we calculated the checklist's overall Cronbach's alpha and the resulting alphas after deleting each item in turn. As a check on the use of domain 1 to classify articles as REC focused, we used Kendall's tau (Noether, 1981) to estimate the correlation between item 1 and each remaining item (2–16).

#### Stage 4: Pretesting

To estimate the IRR in implementation of the final checklist, we prepared scoring instructions for each item (available upon request) and used them to train two coauthors who had not previously participated in data abstraction (M. G., a masters-level data analyst, and L. J. C., a PhD-level researcher). REC-focused articles from stage 2 were selected randomly and coded separately by the two coders until a  $\kappa$  coefficient of greater than 0.90 was attained for the total GAP-REACH score. After an initial training of 1.5 hours, the coders discussed their codes iteratively with the first author and with each other until attaining the desired proficiency ( $n = 8$  articles); total training time was 5 hours.

The two coders then scored a new random sample of 30 articles without stratifying by journal, blinded to each other's results. The sample was selected from among all articles published in 2002 in the seven psychiatric journals that met stage 2 inclusion criteria (*i.e.*, original research with 10 participants) and did not overlap with the abstraction sample in stage 2. Kappa coefficients and 95% confidence limits were obtained for each item and for the overall set of items (16 domains  $\times$  30 articles). To visualize the variability across coders in the pretesting sample of articles, we also plotted the median, quartiles, and whiskers for the 10% and 90% percentiles of the total GAP-REACH score.

## RESULTS

### Stage 1: Development of List of Domains and Data Abstraction Form

Appendix A lists the 16 domains (grouped in five sections) developed by the GAP committee and the representative abstraction questions for each domain. To identify all references to REC factors, the Overall section (domains 1–3) assesses the use of any REC-related term in the title and/or abstract (1) and the article text (2). Domain 3 evaluates whether the REC factors were defined and conceptualized (CDC, 1993; Lee, 2008; Sankar et al., 2007). Under Introduction/Background, we assess whether and why REC factors were considered in the rationale for the study topic and/or design (4; CDC, 1993; Rivara and Finberg, 2001; Sankar et al., 2007). Domains related to Methods (5–12) fall under three subheadings. Under Study Sample (5–7), we include sampling-related domains, in light of federal initiatives to increase reporting of subgroup-level data (NIH, 1994; NIMH, 2001) and investigators' and editors' repeated recommendations to specify how participants' REC characteristics are ascertained (CDC, 1993; Kaplan and Bennett, 2003; Williams, 1994, 1996). The domains under Procedure (8–10) focus on methods to increase the reliability and validity of data collection across REC groups, such as by specifying the language proficiency of study participants (8; IOM, 2009) and disclosing the match (or mismatch) in REC characteristics and/or language fluency between interviewers and participants (which can affect data accuracy; 9–10; Drevdahl et al., 2006). The domains under Instrument Translation and Psychometrics (11–12) assess whether translation methods were described when necessary (11; Williams, 1996) and whether the psychometric adequacy of the instruments was reported for all the REC groups in the study (12; Knight and Hill, 1998; Knight and Zerr, 2010; Williams, 1996). Under Results (13–14), we include domains that indicate an emphasis on REC factors in data analysis. Finally, under Discussion (15–16), we

do the same with respect to data interpretation (Walsh and Ross, 2003). Only information that was explicitly mentioned in the article was assessed as present.

### Stage 2: Data Abstraction

The screening results of titles and abstracts appear in Table 1. This process resulted in the screening of 1205 eligible articles, during a journal-specific time frame of 6 months to 3 years. IRR of the screen results ( $n = 241$ ) was perfect ( $\kappa = 1.0$ ). The screening identified 15 REC-focused articles for each journal except for the *AGP* and the *BP*, which yielded only 12 and 8, respectively, after surveying 3 years of publications. The total sample for data abstraction was therefore  $N = 190$ , including 95 REC-focused and 95 randomly selected non-REC-focused articles.

### Stage 3: Development of the Checklist

**Item Content**—All 16 domains were judged feasible to implement during the abstraction process. Therefore, the abstracted data for all 16 domains were examined to finalize the wording of each item and to determine the criteria for positive scores. Ten items (1–2, 4, 6–10, 12, and 16) were derived almost verbatim from the original wording of the abstraction form questions and scoring options. Items 3, 5, 11, 13, 14, and 15 required additional discussion. Appendix B lists the problems encountered during the abstraction process and the solutions implemented in the final checklist.

**Inapplicability of Items**—Six items (8–13) did not apply to at least some of the articles sampled. As shown in Table 2, the percentage of inapplicability ranged widely by item and journal and was caused by several valid reasons. For example, studies that did not interview individuals (*e.g.*, used claims data) could not be assessed for subjects' language proficiency, match in interviewer-participant characteristics, need for translation, or measurement equivalence.

**Scoring**—We developed a coding scheme that accounts for the extent of inapplicability. Items may be scored yes or no (items 1–7 and 14–16) or yes, no, or not applicable (items 8–13). The GAP-REACH total score is calculated by adding the items scored yes and not applicable, dividing the sum by the total number of items (16), and multiplying by 100%. The total score (range 0%–100%) therefore represents the percentage of items that either were well addressed in the article (coded yes) or did not apply to the research methodology used (coded not applicable).

**Internal Consistency and Correlation Test**—The Cronbach's  $\alpha$  for the total checklist was 0.885. Sequential deletion of individual items did not improve internal consistency ( $\alpha = 0.873$ – $0.883$ ). We also estimated the correlation between item 1 and items 2 to 16, coding not applicable scores as yes scores. All items, except items 8 ( $\tau = 0.12$ ;  $p = 0.10$ ) and 11 ( $\tau = 0.10$ ;  $p = 0.18$ ), were positively correlated with item 1 ( $\tau$  ranging from 0.21 to 0.58;  $p < 0.01$ ). These results confirm that our proxy (based on REC terms in either the title or abstract) successfully classified articles as REC focused.



### Stage 4: Pretesting

Scores and IRR estimates for each item and the total GAPREACH score resulting from our pretesting of 30 articles by two independent coders are presented in Table 3. The time spent coding each article was 30 to 45 minutes. The number and the percentage of the 30 articles meeting each item are presented for each coder. IRR for the items ranged from 0.52 to 1.0, with a  $\kappa$  coefficient of 0.91 for all item-level ratings combined across all articles (16 domains  $\times$  30 articles). IRR estimates for 12 of the 16 items were 0.80 or higher.

Combining yes and not applicable responses, the proportion of items with either of these responses ranged from a low of approximately 7% for both coders on item 16 to a high of 53% (coder 1) to 57% (coder 2) on item 2. The scores for most items were usually less than 20%. Only two items scored greater than 40%: use of at least one REC term in the article text (53%–57%) and description of the sample in terms of REC characteristics (40%–43%). The mean total GAP-REACH score across the two coders was 19.9% (SD, 12.7); a box plot of the total GAP-REACH score percentages by coder is illustrated in Figure 2. The figure indicates, for example, that coder 1 identified 90% of the articles as having a GAP-REACH score lower than 50%.

## DISCUSSION

Systematic reporting of data on REC in psychiatric research enhances the validity and generalizability of all research findings and is essential for understanding and eliminating REC-related disparities in mental health care delivery and outcomes (Drevdahl et al., 2006; IOM, 2009). This article describes the process by which the GAP Cultural Committee developed a checklist to operationalize and quantify the level of comprehensiveness of REC factor reporting in psychiatric research publications.

### Development of the Checklist

We used expert consensus, empirical review of publication content, and item-based reliability analysis to develop the GAP-REACH checklist. Our coding scheme evolved to account for the possible inapplicability of checklist items to certain methods (*e.g.*, analysis of claims data). This proved essential because journals differed significantly in the proportion of item inapplicability as a result of legitimate methodological differences (see Table 2). The GAP-REACH checklist thus allows items to be bypassed if these do not apply to the methodology used; only poorly addressed items are identified as deficient and lower the total GAP-REACH score. We considered the alternative of excluding inapplicable items from the coding scheme (*i.e.*, basing the final score only on positive and negative items). However, this resulted paradoxically in the finding that the negative items were then weighted more heavily in the final score, thereby lowering it. Effectively, this would have penalized authors for using research methods that were independent of REC factors (*i.e.*, not applicable in checklist terms), which was not our intent. Using the final coding scheme, the resulting checklist had excellent internal consistency ( $\alpha = 0.885$ ).

In developing the checklist domains, we chose to assess reporting of data on REC separately from reporting of socioeconomic variables, despite critiques that this separation could help

misattribute the effects of socioeconomic status (SES) to REC factors (CDC, 1993; Kaplan and Bennett, 2003; Rivara and Finberg, 2001). Three reasons justify this approach. First, in clinical care and policy making, disparities by race/ethnicity and cultural background remain important as markers of inadequate care, of differences in clinical presentation, or of variation in diagnostic assessment and treatment response, independently of their interaction with other factors, such as income or education (Krieger et al., 1993; Williams, 1996). It is important to identify and track these REC-related disparities, regardless of whether these are ultimately caused by system-level (*e.g.*, rates of noninsurance), provider-level (*e.g.*, low cultural competency in care delivery), or patient-level factors (*e.g.*, cultural variability in symptoms). Second, the research practice of adjusting for SES as a way of isolating the independent effect of REC factors usually disregards the impact of other sociohistorical factors such as racism or segregation that may confound or moderate the effect of SES variables across REC groups. Simple adjustment by SES does not take these factors into account, highlighting the need to continue to assess race/ethnicity in health research (Baker et al., 2005; Dixon et al., 2011; Kessler and Neighbors, 1986; Williams, 1996). Third, race/ethnicity variables often act as proxies for explicitly cultural factors, which are not reducible to SES variables and frequently go unmeasured. People from similar educational or income backgrounds can differ markedly in their experience of mental illness, their explanatory models, and their views of mental health care as a result of frankly cultural factors, such as religion, alternative illness expressions and healing traditions, or local health care practices (Kleinman, 1988). The GAP-REACH checklist reflects the view that race/ethnicity should still be collected in health research, while also endorsing the need to go beyond these simpler indices to direct measurement of cultural elements (Dixon et al., 2011; Krieger et al., 1993; Mir et al., 2012; Williams, 1996).

### Interrater Reliability

Item-based analysis of the final checklist among our two masters- and doctoral-level coders revealed excellent IRR of the set of all items (0.91) and moderate to excellent IRR reliability of individual items (0.52–1.0). Only four items (5, 7, 13, and 15) yielded  $\kappa$  coefficients of less than 0.80. The item with the lowest IRR is discussed below.

Item 5 ( $\kappa = 0.52$ ) assesses whether a study's sampling procedure has taken into account the distribution of REC characteristics among the participants. This item requires judgment on the part of the coder because it attempts to ascertain the investigators' intent with respect to the role of REC factors in sampling, which is often not explicitly reported. This determination was particularly difficult to make for studies conducted in populations that were presented as racially and ethnically homogenous. If a study from Japan, for example, included only Japanese subjects, was this because the article intended to report something particular to this population (*e.g.*, the prevalence of certain genetic polymorphisms as compared with US and European findings)? Or was it simply a sample of convenience in a Japanese setting and therefore without specific attention to REC factors? This crucial distinction led to disagreement in 3 of 30 cases.

As expected, items that required less judgment by the coders (*e.g.*, those that hinged on the presence or absence of specific words, such as items 1 and 2) had the highest IRR (Oxman

et al., 1991). However, it should be noted that several items that required substantial judgment in coding still achieved excellent IRR (*e.g.*, items 4, 8, 12). Moreover, coding mismatches occurred in very few articles for each item (never in >10% of articles), suggesting that the items are feasible to apply and that improvements in IRR may result from more detailed scoring instructions or from further training than was provided, especially for items that require more judgment. Some uses of the checklist would not be affected by training effects. For example, authors using the GAP-REACH during the design phase of their study could easily complete the items that require judgment of the investigators' intent.

### Potential Checklist Implications

At this early stage of development, the purpose of the GAPREACH checklist is purely descriptive. Further work is needed to develop it as a guideline for minimal standards in assessing publication. For example, description of the REC characteristics of the sample (item 6) or of the measurement equivalence of instruments across all REC groups in the study (item 12) are candidate items that may be required of every research article that uses instruments to assess diverse study populations. Previous efforts to improve REC reporting have had limited impact (Mak et al., 2007; Walsh and Ross, 2003), possibly because guidelines were not successfully operationalized as a list of domains. Even when federal mandates have led to improved gathering of REC-related data, this has not necessarily translated into more systematic reporting. For example, a survey of authors of pediatric articles published in 1999–2000 found that only half of the researchers who collected REC-related data because of the 1993 NIH requirement subsequently reported these findings (Walsh and Ross, 2003). In mental health, the effect of the 2001 NIMH initiative on quality of REC reporting and extent of subgroup analyses was still limited 3 years later (Mak et al., 2007). Further development of the GAP-REACH checklist into a set of editorial recommendations could highlight these discrepancies and help guide reporting practice.

In its current form, the GAP-REACH can be used before or after publication by a broad range of stakeholders in the research process. Users may emphasize individual items or the total GAP-REACH score. Investigators will likely benefit from focusing on separate items across all stages of the research process to check the comprehensiveness of REC factor reporting. This approach may be especially helpful for studies that aim to improve racial/ethnic minority and multicultural health and the elimination of disparities in mental health care. Journal reviewers and editors may use the domains to guide their reviews and the total score to track the journal's use of REC factors over time and to develop editorial policies. Clinical readers may use the checklist to differentiate REC-related issues in the delivery of interventions and in their assessments to guide treatment planning so that culturally appropriate care may help to reduce disparities (Ruiz and Primm, 2010). The total score may also help clinicians to assess whether the article findings were obtained with sufficient attention to REC factors to warrant application to the populations they serve. Funding agencies, government and private health care administrators, as well as policy makers and advocacy groups can use the total score to foster initiatives designed to improve the generalizability of mental health research across diverse populations and to evaluate their impact.

Active endorsement by multiple stakeholders (e.g., editors, mental health organizations) is necessary for the GAP-REACH to successfully impact psychiatric research. A “guidance checklist” recently developed in the United Kingdom to improve the use of race/ethnicity variables in general health research was not widely adopted by journal reviewers (Salway et al., 2011a, 2011b). This 21-item checklist shares common features with the GAP-REACH (e.g., focus on sampling strategy). However, it also differs in several ways: it is longer, it is less focused on psychometric adequacy, it does not include culture in its purview, its items require greater interpretation by the user, it does not result in a total score that can be compared across articles and journals, and it pays very limited attention to the applicability of the checklist across diverse types of research articles. Future work on the GAP-REACH and the UK checklist must focus on how to enhance their utility to facilitate their adoption by various stakeholders in the research process.

### Findings From Screening and Pretesting

In addition to assessing the comprehensiveness of REC factor reporting within individual publications, researchers may use the GAP-REACH to evaluate the overall status of the aggregate reporting of REC factors in more overarching domains of psychiatric research. For example, individual checklist scores may be averaged across groups of articles to characterize the use of REC factors at the journal level or for the field as a whole. The data from stages 2 and 4 demonstrate the limitations of the field at the time of the 2001 NIMH initiative to improve reporting of REC factors. In stage 2, screening of the seven psychiatric journals with the first checklist item revealed that in two widely read publications, the *BP* and the *AGP*, REC terms were included in only 2% to 4% of the article titles or abstracts during 2000–2002. Although the relevance of REC factors to these articles was not assessed by the checklist, these findings still raise questions about missed opportunities to include REC factors in these publications.

The data from stage 4 of a random sample of articles published in 2002 are more revealing because it is based on the full checklist. Table 3 shows how infrequently REC factors were reported in 2002, suggesting persistent inattention in psychiatric publications. Application of the GAP-REACH checklist can act as a corrective.

### Limitations

The final checklist aggregates variables related to REC into a single construct labeled “REC”; this may unintentionally obscure or elide important differences across these concepts. It would have been preferable to assess each concept separately, but this would have resulted in a lengthy instrument, which would likely reduce the feasibility of its use. The expert consensus process may have missed elements of REC reporting or overemphasized some at the expense of others. Domain development was based on a limited number of journals, possibly resulting in reduced applicability of the checklist to other journals or subspecialties within psychiatry. Our proxy measure for selecting REC-focused articles based on item 1 may have missed some articles that could have helped us refine domain content and item scoring. However, the correlation checks between item 1 and the rest of the checklist were significant for all but two items, suggesting that the articles selected were adequate to the task. Development of scoring options was based on expert

judgment as to what constituted a fair response for each domain, and different choices would affect scoring results. Our rationale was documented in each case but ultimately was based on expert opinion. The total GAP-REACH score is a simple proportion, which attributes equivalent importance across items; it is arguable that some items should contribute more than others to the total score. Further research on the checklist is needed to clarify this and refine the most informative and valid scoring method. Our pretesting sample was small, limiting our ability to report on cross-journal differences. In this pretest, we used MA-level and PhD-level coders, so we cannot be sure that coders with less research experience would obtain the same results. Finally, we are confined to the information provided by the authors. Investigators may have obtained data that they did not report but their intent study design; still, the checklist results represent what was reported to the field.

The absence of a designated cut point for acceptable inclusion of REC variables in a research article may be regarded as a limitation. After careful consideration, we concluded that assigning a cut point for an “adequate,” “good,” or “poor” score was beyond the scope of this article. Such a cut point requires further consideration that would depend in part on the topic of the article or the type of research study and the extent to which previous research had shown REC factors to be relevant.

## CONCLUSIONS

The GAP-REACH checklist was developed to evaluate the scope of reporting of REC factors in psychiatric research publications. Its widespread implementation is encouraged, particularly by investigators, journal editors, and reviewers. The format and content of the final checklist may also be adapted to other clinical specialties in which there is not yet adequate systematic application of REC factors in research. The next step will be to apply the checklist to a representative sample of psychiatric publications from 2003 to 2012—subsequent to the time frame used to develop the checklist—to track changes in REC reporting over time. Only by systematic study and reporting of REC factors will our field generate the empirical base critical to the elimination of disparities in health and health care affecting diverse racial, ethnic, and cultural populations.

## ACKNOWLEDGMENTS

The authors thank Ivan Balán, Michelle Bell, Melissa Rosario, Naelys Díaz, Dana Alonzo, Spencer Cruz-Katz, Henian Chen, Sewon Kim, Mandy Garber, Sapana Patel, Phil Saccone, Luz Marte, Junius J. Gonzales, Laurence Kirmayer, John Markowitz, and Naudia Pickens for their substantial help.

This study was supported by grants R21 MH066388, R34 MH073087, and R01 MH077226 from the National Institute of Mental Health and a grant award from the Brain and Behavior Research Foundation to Dr. Lewis-Fernández and institutional funds from the New York State Psychiatric Institute.

## REFERENCES

- Al Faleh K, Al-Omran M. Reporting and methodological quality of Cochrane Neonatal Review Group systematic reviews. *BMC Pediatr*. 2009; 9:38. [PubMed: 19534780]
- American Academy of Pediatrics Committee on Pediatric Research. Race/ethnicity, gender, socioeconomic status—Research exploring their effects on child health: A subject review. *Pediatrics*. 2000; 105:1349–1351. [PubMed: 10835080]

- Baker EA, Metzler MM, Galea S. Addressing social determinants of health inequities: Learning from doing. *Am J Public Health*. 2005; 95:553–555. [PubMed: 15798106]
- Barley R, Salway S. Ethical practice and scientific standards in researching ethnicity: Methodological challenges of conducting a literature review. 2008 Retrieved from <http://research.shu.ac.uk/ethics-ethnicity/docs/newdocs/Methodologicalchallengesofconductingaliteraturereview>.
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA*. 1996; 276:637–639. [PubMed: 8773637]
- Bell C, Williamson J. Articles on special populations published in *Psychiatric Services* between 1950 and 1999. *Psychiatr Serv*. 1999; 53:419–424. [PubMed: 11919354]
- Bhopal R, Kohli H. Editors' practice and views on terminology in ethnicity and health research. *Ethn Health*. 1997; 2:223–227. [PubMed: 9426986]
- Centers for Disease Control and Prevention. Use of race and ethnicity in public health surveillance: Summary of the CDC/ATSDR workshop. *MMWR Morb Mortal Wkly Rep*. 1993; 42:1–16. [PubMed: 8418395]
- Comstock RD, Castillo EM, Lindsay SP. Four-year review of the use of race and ethnicity in epidemiologic and public health research. *Am J Epidemiol*. 2004; 159:611–619. [PubMed: 15003966]
- Department of Health and Human Services. HHS policy for improving race and ethnicity data. Washington, DC: Department of Health and Human Services; 1997. Retrieved from <http://aspe.hhs.gov/dataacncl/inclusn.htm>.
- Department of Health and Human Services. Affordable Care Act to improve data collection, reduce health disparities. Washington, DC: Department of Health and Human Services; 2011. Retrieved from <http://www.hhs.gov/news/press/2011pres/06/20110629a.html>. [Accessed on June 15, 2012]
- Dixon L, Lewis-Fernández R, Goldman H, Interian A, Michaels A, Riley M. Adherence disparities in mental health: Opportunities and challenges. *J Nerv Ment Dis*. 2011; 199:815–820. [PubMed: 21964279]
- Drevdahl DJ, Philips DA, Taylor JY. Uncontested categories: The use of race and ethnicity variables in nursing research. *Nurs Inq*. 2006; 13:52–63. [PubMed: 16494667]
- Imada T, Schiavo RS. The use of ethnic minority populations in published psychological research, 1990–1999. *J Psychol*. 2005; 139:389–400. [PubMed: 16285210]
- Institute of Medicine. Race, ethnicity, and language data: Standardization for health care quality improvement. 2009 Retrieved from <http://nationalacademies.org/includes/data.pdf>.
- International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication. Philadelphia: International Committee of Medical Journal Editors; 2010. Selection and description of participants. (Section IV.A.6.a). Retrieved from <http://www.icmje.org>. [Accessed on January 5, 2011]
- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds JM, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Control Clin Trials*. 1996; 17:1–12. [PubMed: 8721797]
- Kahn J. Getting the numbers right: Statistical mischief and racial profiling in heart failure research. *Perspect Biol Med*. 2003; 46:473–483. [PubMed: 14593217]
- Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication. *JAMA*. 2003; 289:2709–2716. [PubMed: 12771118]
- Kessler RC, Neighbors HW. A new perspective on the relationship among race, social class, and psychological distress. *J Health Soc Behav*. 1986; 27:107–115. [PubMed: 3734380]
- Kleinman, A. Rethinking psychiatry: From cultural category to personal experience. New York: Free Press; 1988.
- Knight, GP.; Hill, NE. Measurement equivalence in research involving minority adolescents. In: McLoyd, VC.; Steinberg, L., editors. *Studying minority adolescents: Conceptual, methodological, and theoretical issues*. Mahwah, NJ: Lawrence Erlbaum Associates; 1998.
- Knight GP, Zerr AA. Introduction to the special section: Measurement equivalence in child development research. *Child Dev Perspect*. 2010; 4:1–4.

- Kocsis JH, Gerber AJ, Milrod B, Roose SP, Barber J, Thase ME, Perkins P, Leon AC. A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Compr Psychiatry*. 2010; 51:319–324. [PubMed: 20399343]
- Krieger N, Rowley DL, Herman AA, Avery B, Phillips MT. Racism, sexism, and social class: Implications for studies of health, disease, and well-being. *Am J Prev Med*. 1993; 9:82–122. [PubMed: 8123288]
- Lee C. “Race” and “ethnicity” in biomedical research: How do scientists construct and explain differences in health? *Soc Sci Med*. 2008; 68:1183–1190. [PubMed: 19185964]
- Mak W, Law R, Alvidrez J, Pérez-Stable E. Gender and ethnic diversity in NIMH-funded clinical trials: Review of a decade of published research. *Adm Policy Ment Health*. 2007; 34:497–503. [PubMed: 17690976]
- Mir G, Salway S, Kai J, Karlsen S, Bhopal R, Ellison GTH, Sheikh A. Principles for research on ethnicity and health: The Leeds Consensus Statement. *Eur J Public Health*. 2012; 22:1–7.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUORUM statement. *Lancet*. 1999; 354:1896–1900. [PubMed: 10584742]
- Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *JAMA*. 2001; 285:1992–1995. [PubMed: 11308436]
- Moher D, Liberati A, Tetzlaff J, Altman DG, for The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *BMJ*. 2009; 339:b2535. [PubMed: 19622551]
- Moher D, Weeks L, Ocampo M, Seely D, Sampson M, Altman DG, Schulz KF, Miller D, Simera I, Grimshaw J, Hoey J. Describing reporting guidelines for health research: A systematic review. *J Clin Epidemiol*. 2011; 64:718–742. [published online ahead of print]. [PubMed: 21216130]
- National Institutes of Health. Revitalization Act of 1993 subtitle B—Clinical research equity regarding women and minorities. Bethesda, MD: National Institutes of Health; 1993. Retrieved from <http://orwh.od.nih.gov/inclusion/revitalization.pdf>. [Accessed on January 3, 2011]
- National Institutes of Health. NIH guidelines on the inclusion of women and minorities as subjects in clinical research. *Federal Register*. 1994; 14:508. 59, (Document no. 94–5435). Retrieved from <http://www.grants.nih.gov/grants/guide/notice-files/not94-100.html>.
- National Institutes of Health. NIH policy on reporting race and ethnicity data: Subjects in clinical research. Bethesda, MD: National Institutes of Health; 2001. Retrieved from <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-01-053.html>. [Accessed on January 3, 2011]
- National Institute of Mental Health. Five-year strategic plan for reducing health disparities. Rockville, MD: National Institute of Mental Health; 2001. Retrieved from <http://www.nimh.nih.gov/about/strategic-planning-reports/nimh-five-year-strategic-plan-for-reducing-health-disparities.pdf>. [Accessed on January 3, 2011]
- National Partnership for Action to End Health Disparities. National stakeholder strategy for achieving health equity. Washington, DC: National Partnership for Action; 2011. Retrieved from <http://www.minorityhealth.hhs.gov/npa/templates/content.aspx?vl=1&lvlid=33&ID=286>. [Accessed on November 27, 2012]
- Noether GE. Why Kendall Tau??. *Teach Stat*. 1981; 3:41–43.
- Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, Streiner DL. Agreement among reviewers of review articles. *J Clin Epidemiol*. 1991; 44:91–98. [PubMed: 1824710]
- Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the CONSORT checklist improve the quality of reports of randomised clinical trials?: A systematic review. *Med J Aust*. 2006; 185:263–267. [PubMed: 16948622]
- Rivara FP, Finberg L. Use of the terms race and ethnicity. *Arch Pediatr Adolesc Med*. 2001; 155:119. [PubMed: 11177083]
- Ruiz, P.; Primm, A., editors. Disparities in psychiatric care: Clinical and cross-cultural perspectives. Baltimore: Lippincott Williams & Wilkins; 2010.

- Salway S, Barley R, Allmark P, Gerrish K, Higginbottom G, Johnson MRD, Ellison GTH. Enhancing the quality of published research on ethnicity and health: Is journal guidance feasible and useful? *Divers Health Care*. 2011a; 8:155–165.
- Salway S, Barley R, Allmark P, Gerrish K, Higginbottom G, Ellison G. Ethnic diversity and inequality: Ethical and scientific rigour in social research. 2011b Retrieved from <http://www.jrf.org.uk/sites/files/jrf/ethnicity-social-policy-research-full.pdf>.
- Sankar P, Cho MK, Mountain J. Race and ethnicity in genetic research. *Am J Med Genet A*. 2007; 143A:961–970. [PubMed: 17345638]
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA*. 2000; 283:2008–2012. [PubMed: 10789670]
- Walsh C, Ross LF. Whether and why pediatrics researchers report race and ethnicity. *Arch Pediatr Adolesc Med*. 2003; 157:671–675. [PubMed: 12860789]
- Williams DR. The concept of race. *Health Services Research: 1966 to 1990*. *Health Serv Res*. 1994; 29:261–274.
- Williams DR. Race/ethnicity and socioeconomic status: Measurement and methodological issues. *Int J Health Serv*. 1996; 26:483–505. [PubMed: 8840198]

## Appendix A

### DOMAINS AND REPRESENTATIVE QUESTIONS IN THE SCREENING FORM (DOMAIN 1) AND ABSTRACTION FORM (2–16) USED TO DEVELOP THE GAP-REACH CHECKLIST

Domain Number	Heading	Domain	Representative Question(s)
1	Overall	Include REC factors in the title and/or abstract	Are the following terms (or related terms) used in the article title or abstract? yes/no (Race, ethnicity, culture, Hispanic/Latino, Black/African American, White/Caucasian, Asian American, American Indian, Ethno-national group (Specify: _____))
2		Include REC factors in the article text	Note verbatim any REC-related term(s) used in the article text: _____
3		Define REC factor(s)	Do the authors provide a definition or conceptualization of REC factors? yes/no Note term(s) used: _____ and summarize definition briefly: _____
4	Introduction/ Background	Discuss the role of REC factors in the rationale for the study	Do the authors discuss the rationale for the study topic or study design in terms of REC factors? mentioned/not mentioned Note category(ies) used: _____ ( <i>e.g.</i> , race, ethnicity)
5	Methods Study sample	Include REC factors in sampling procedure	Note overall characteristics of sampling method: <ul style="list-style-type: none"> <li>a. No attention paid to REC</li> <li>b. Single REC group</li> <li>c. Sample stratified by REC</li> <li>d. Oversampling of REC groups</li> <li>e. Other strategy: _____</li> </ul>
6		Describe REC characteristics of the sample	Do the authors describe their sample in terms of REC variables? yes/no Note (verbatim) the categories used by the authors to describe the sample in terms of REC factors: _____



Domain Number	Heading	Domain	Representative Question(s)
7		Describe how participants' REC characteristics were ascertained	Do the authors indicate how these groups were assessed in terms of REC categories? <ul style="list-style-type: none"> <li>a. Not specified</li> <li>b. Self-report</li> <li>c. Place of birth</li> <li>d. Interviewer</li> <li>e. Other: _____</li> </ul>
8	Procedure	Specify proficiency of participants in the language(s) of the study	Do the authors mention whether a specific language proficiency was a requirement for study entry? Yes/no/not applicable due to: _____ Did the authors specify a method to determine language proficiency? yes/no
9		Mention match or mismatch of interviewers' and participants' REC characteristics	Do the authors mention the relevance of the REC characteristics of the interviewers vis-à-vis the REC characteristics of the participants? yes/no/not applicable due to: _____
10		Mention match or mismatch of interviewers' and participants' language characteristics	Do the authors mention the relevance of the language fluency of the interviewers vis-à-vis the language fluency of the participants? Yes/no/not applicable due to: _____
11	Instrument translation and psychometrics	Report on translation of instruments when these were not developed in the language of the study population	Were the study instruments originally created in the language of the study population? yes/no If no: Do the authors mention whether the study instruments were translated into the language(s) of the study population? yes/no
12		Discuss measurement equivalence of instruments for all REC groups included	Do the authors assess the measurement equivalence of the instruments for all the defined REC populations in the study? yes/no/not applicable due to: _____
13	Results	Test effect of REC factors on study outcome(s)	Do the authors test the bivariate association between REC variables and any outcome variables? yes/no/not applicable due to: _____
14		Include REC factors in data analysis	Choose one role for REC categories in the overall analysis: <ul style="list-style-type: none"> <li>a. Covariate but not a main effect variable (<i>e.g.</i>, relationship between non-REC independent and dependent variables adjusted for REC in multiple regression)</li> <li>b. As a main effect variable (<i>e.g.</i>, relationship between REC variables and outcome variables adjusted for non-REC variables in multiple regression)</li> <li>c. Not applicable due to: _____</li> <li>d. Not done</li> </ul> Describe the exact role of the REC variables in the overall analysis: _____

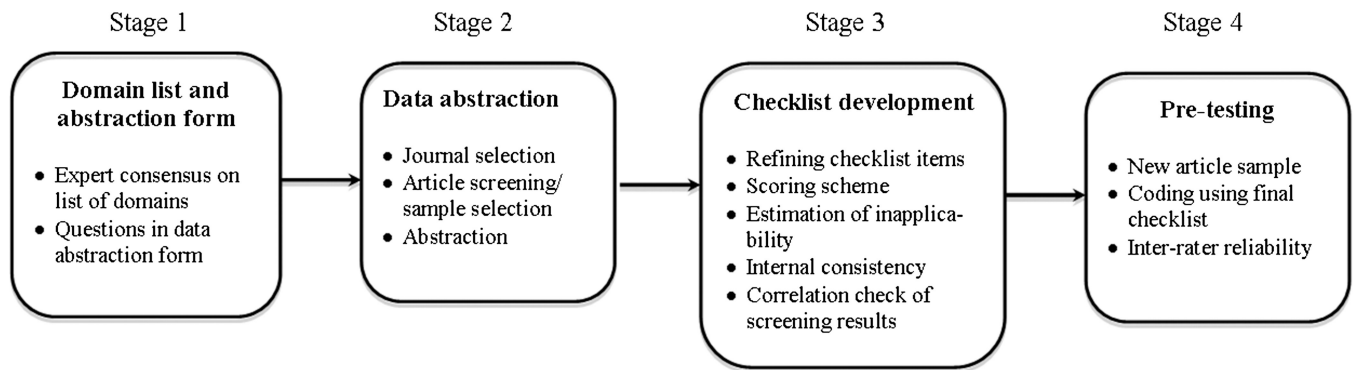
Domain Number	Heading	Domain	Representative Question(s)
15	Discussion	Emphasize REC factors in the interpretation of results	Do the authors refer to any REC factors in the interpretation of their results? a. Not at all b. Somewhat c. Moderately d. A great deal
16		Include REC factors in the discussion of study limitations	Do the authors discuss study limitations in terms of REC factors? yes/no/not applicable due to: _____

## Appendix B

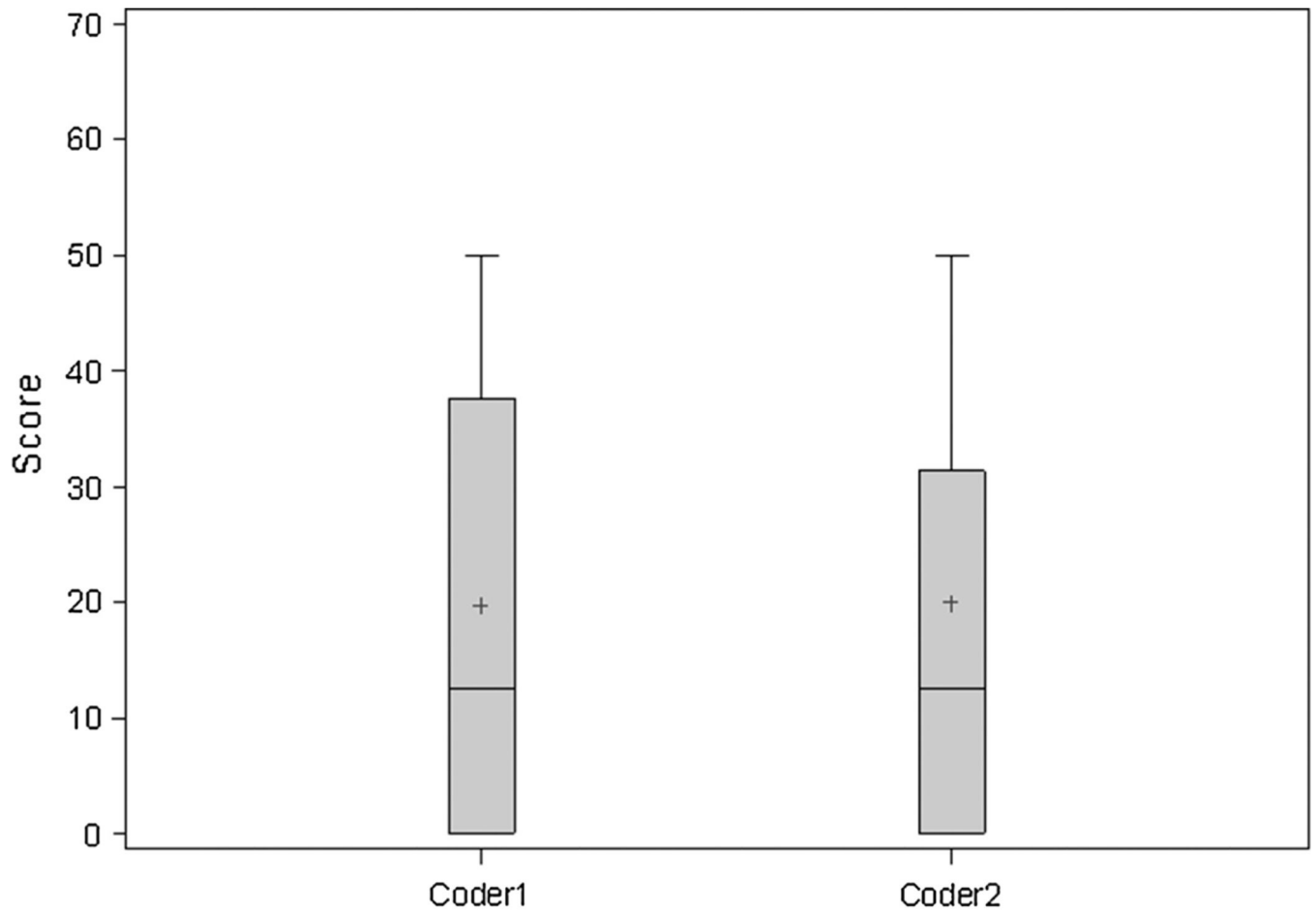
### PROBLEMS ENCOUNTERED DURING ABSTRACTION PROCESS (N = 190) AND SOLUTIONS IMPLEMENTED IN FINAL CHECKLIST

Item	Domain	Problem Encountered During Abstraction	Checklist Solution
3	Define REC factors	Definitions and conceptualizations of REC factors often not provided or done in rudimentary way	Counted any definition of REC terms as positive, even if vague or basic ( <i>e.g.</i> , “we used the 2000 Census categories”).
5	Include REC factors in sampling procedure	Coded as four complex questions in abstraction form	Condensed into single item. Coded “positive” only articles that attended specifically to REC factors in sampling ( <i>e.g.</i> , justified sampling a single REC group, stratified by REC categories). Other articles coded “negative” ( <i>e.g.</i> , article reported samples of convenience without discussing relevance of REC factors).
11	Describe translation methods	Coded as three complex questions in abstraction form Studies conducted in English and published in English did not report translation methods although language fluency of all participants were not always reported	Condensed into single item. Coded “not applicable” articles that: <ul style="list-style-type: none"> <li>Used no instruments (<i>e.g.</i>, claims data)</li> <li>Reported use of instruments that were created in the language of participants’ required proficiency</li> <li>Explicitly required English proficiency of all participants because all articles were published in English and thus were assumed to use English-language instruments</li> </ul> Coded “negative” articles that did not report instrument translation and also: <ul style="list-style-type: none"> <li>Did not report participants’ language proficiency</li> <li>Required non-English-language proficiency but did not report match between this proficiency and the instruments used</li> </ul> Coded “positive” articles that reported instrument translation, whether language proficiency was required.

Item	Domain	Problem Encountered During Abstraction	Checklist Solution
13	Test REC effect on outcome	Need to distinguish articles that intentionally sampled a single REC group (and therefore did not include REC terms directly in analyses) and articles that could have conducted REC-based analyses but did not ( <i>e.g.</i> , included multiple REC groups).	Coded as “not applicable” studies that sampled a single REC group with the intent of studying that particular population.
14	Include REC factors in data analysis	Coded as two questions in abstraction form Same issue as above for item 13.	Condensed into single item Coded “positive” any inclusion of REC factors in analyses, including articles that sampled a single REC group with the intent of studying that particular population.
15	Include REC factors in study limitations	Abstraction form question had four ordinal response options.	Dichotomized response options Coded “not at all” and “somewhat” as “negative” Coded “moderate” and “a great deal” as “positive”



**FIGURE 1.** Stages and steps in the development of the GAP-REACH checklist.



Note: Outer values are 10<sup>th</sup> and 90<sup>th</sup> percentiles; 10<sup>th</sup> and 25<sup>th</sup> percentiles are 0 for both coders. Mean is denoted by + symbol.

**FIGURE 2.**

Distribution of GAP-REACH scores in the pretesting sample ( $n = 30$ ), by coder.

**TABLE 1**

Number and Percentage of REC-Focused Articles per Journal by Quarter (January 2000 to December 2002)

Journal	2000				2001				2002				Total Screened (N = 1205)	REC Focused, <i>a</i> %		
	1	2	3	4	1	2	3	4	1	2	3	4			Total REC Focused (n = 95)	
<i>AGP</i>	0	1	1	0	2	2	2	1	1	2	0	0	2	12	286	3.9
<i>AJP</i>	—	—	—	—	—	—	—	—	—	—	8	7	15	125	11.3	
<i>BP</i>	0	0	0	0	2	0	2	0	0	0	0	1	3	8	305	2.3
<i>JAACAP</i>	—	—	—	—	—	—	—	—	—	5	1	5	4	15	115	12.3
<i>JCP</i>	—	—	—	—	—	2	2	0	5	2	3	1	15	198	7.1	
<i>JNMD</i>	—	—	—	—	—	—	—	—	—	4	7	3	1	15	76	18.7
<i>PS</i>	—	—	—	—	—	—	—	—	1	1	2	8	3	15	100	14.1

<sup>a</sup>Rate based on the uniformly minimum variance unbiased estimator.

**TABLE 2**  
 Proportion and Reasons of Inapplicability by Journal for the GAP-REACH Checklist Items 8 to 13

Item No.	Checklist domain	Journal										Reasons <sup>d</sup>
		AGP	AJP	BP	JCP	JNMD	JAACAP	FS	$\chi^2$	p		
8	Language proficiency	12.5%	16.7%	0.0%	6.7%	3.3%	10.0%	23.3%	24.2	<0.05	Administrative data set: 95% Chart abstraction: 14%	
9	REC relevance	16.7%	20.0%	0.0%	10.0%	3.3%	10.0%	33.3%	23.0	<0.05	Administrative data set: 78% Self-report questionnaire: 19% Chart abstraction: 15%	
10	Language relevance	16.7%	20.0%	0.0%	10.0%	3.3%	10.0%	33.3%	25.2	<0.05	Administrative data set: 78% Self-report questionnaire: 19% Chart abstraction: 15%	
11	Translation	75.0%	63.3%	56.3%	76.7%	53.3%	66.7%	80.0%	15.5	0.21	No need for translation: 79% Administrative data set: 16% Chart abstraction: 5% Unstructured interview: 2%	
12	Measurement equivalence	12.5%	20.0%	0.0%	13.3%	6.7%	13.3%	30.0%	24.9	<0.05	Administrative data set: 75% Chart abstraction: 21% Unstructured interview: 14%	
13	Test REC effect on	4.2%	16.7%	25.0%	10.0%	10.0%	13.3%	6.7%	14.1	0.29	Single REC group: 86% Stratification or matching by REC: 14%	

<sup>d</sup> Responses total to more than 100% because of multiple reasons cited for some articles.

**TABLE 3**  
 Scores and IRR From Pretesting of the GAP-REACH Checklist and the GAP-REACH Score Assessing the Use of REC-Related Factors in a Random Sample of Psychiatric Research Articles Published in 2002

Item	Coder 1		Coder 2		κ	95% Confidence Limit
	n <sup>a</sup>	%	n <sup>a</sup>	%		
<b>Overall REC use</b>						
1. At least one REC term used in the title or abstract	4	13.3	4	13.3	1.0	1.0–1.0
2. At least one REC term used in the article text	16	53.3	17	56.7	0.93	0.80–1.0
3. Definition/conceptualization of REC term(s) provided	3	10.0	3	10.0	1.0	1.0–1.0
<b>Introduction/Background</b>						
4. Rationale for study question or design discussed in terms of REC factors	4	13.3	4	13.3	1.0	1.0–1.0
<b>Methods</b>						
5. REC included in sampling procedure	3	10.0	4	13.3	0.52	0.04–0.99
6. Sample described in terms of REC characteristics	13	43.3	12	40.0	0.93	0.80–1.0
7. Method described for assessing REC characteristics of participants	4	13.3	7	23.3	0.67	0.34–1.0
8. Language proficiency of participants specified	7	23.3	8	26.7	0.92	0.76–1.0
9. Relevance of REC characteristics of interviewers and participants discussed	4	13.3	4	13.3	1.0	1.0–1.0
10. Relevance of language characteristics of interviewers and participants discussed	5	16.7	5	16.7	1.0	1.0–1.0
11. Language match between participants and instruments is appropriately described	5	16.7	5	16.7	1.0	1.0–1.0
12. Measurement equivalence of instruments described for all REC groups in the study	6	20.0	5	16.7	0.89	0.69–1.0
<b>Results</b>						
13. Effect of REC factors on study outcome(s) tested	10	33.3	7	23.3	0.78	0.55–1.0
14. REC factors included in data analysis	7	23.3	6	20.0	0.90	0.71–1.0
<b>Discussion</b>						
15. REC emphasized in data interpretation	2	6.7	3	10.0	0.78	0.37–1.0
16. Study limitations discussed in REC terms	2	6.7	2	6.7	1.0	1.0–1.0
Sum of items	95		96		0.91	0.86–0.95.
Average GAP-REACH score (30 articles)		19.8		20.0		

Data are presented as number and percentage of articles (n = 30) meeting each item.

Coder 1: item 8: yes (n = 3), NA (n = 4); item 9: yes (n = 0), NA (n = 4); item 10: yes (n = 1), NA (n = 4); item 11: yes (n = 1), NA (n = 4); item 12: yes (n = 2), NA (n = 4); item 13: yes (n = 5), NA (n = 5).  
 Coder 2: item 8: yes (n = 4), NA (n = 4); item 9: yes (n = 0), NA (n = 4); item 10: yes (n = 1), NA (n = 4); item 11: yes (n = 1), NA (n = 4); item 12: yes (n = 1), NA (n = 4); item 13: yes (n = 5), NA (n = 2).



NA indicates not applicable.

<sup>a</sup>Includes articles coded yes for items 1 to 7 and 14 to 16 and for the GAP-REACH score. For items 8 to 13, includes articles coded yes plus articles coded not applicable.