

An Assessment of the Quality of Competence Restoration Research

Philip J. Candilis, MD, and George F. Parker, MD

A systematic review of the literature on restoration of competence to stand trial identified a predominance of retrospective case studies using descriptive and correlational statistics. Guided by National Institutes of Health (NIH) quality metrics and emphasizing study design, sample size, and statistical methods, the authors categorized a large majority of studies as fair in quality, underscoring the need for controlled designs, larger representative samples, and more sophisticated statistical analyses. Implications for the state of forensic research include the need to use large databases within jurisdictions and the importance of reliable methods that can be applied across jurisdictions and aggregated for meta-analysis. More sophisticated research methods can be advanced in forensic fellowship training where coordinated projects and curricula can encourage systematic approaches to forensic research.

J Am Acad Psychiatry Law 52:153–60, 2024. DOI:10.29158/JAAPL.240008-24

Key words: adjudicative competence; competence restoration; systematic review; research quality

The conduct and quality of empirical research in forensic psychiatry is affected by specific ethics and procedural constraints. The randomized controlled trial favored by clinical medicine is unavailable for many of the interventions sought by defendants and courts, from diversion, dismissal, or probation, to incarceration, competence assessment, or restoration. Outcomes are determined by specific factfinders, and by the legal strategies chosen by defendants in specific circumstances. Even established interventions like competence restoration strategies (i.e., group or individual education, medication classes, or outpatient, inpatient, and jail-based programming) are affected by numerous factors: the requirements of a case, the rules and resources of a jurisdiction, and the assessments of individual judges, testifying experts, defendants, and their attorneys. It is a significant challenge for those in forensic mental health to identify and control target samples, interventions, and settings.

Research protections may also limit exploration of specific interventions among detained groups. The term “prisoner,” for example, as used by the U.S. Department of Health and Human Services (DHHS) and its Office of Human Research Protections (OHRP) extends to persons under civil detention or awaiting trial and sentencing.¹ Research may not be approved if it does not follow specific oversight requirements like the presence of a prisoner representative on the institutional review board. Research may not fall into approved categories such as improving the conditions affecting participants or studying the effects of incarceration. Use of control groups that may not benefit research participants is particularly sensitive and requires the DHHS Secretary’s explicit review.²

Forensic training itself has traditionally focused on conducting forensic evaluations, preparing narrative reports, and studying the landmark court decisions that shaped the field. Research projects during and after training consequently emphasize the clinical aspects of the specialty, leading to a predominance of case series in the literature, typically using descriptive statistics.^{3,4}

This context has specific implications for forensic research, limiting its opportunities, depth, and reach. The quality of research accepted in court is decided by lay judges scrutinizing the contributions of individual experts, and is governed by *Daubert*⁵ and *Frye*⁶ criteria that, though legal, draw on medical expertise on

Published online April 30, 2024.

Dr. Candilis is Medical Director of Saint Elizabeths Hospital, Washington, D.C., and a Professor of Psychiatry, George Washington University School of Medicine and Health Sciences, Washington, DC. Dr. Parker is Professor of Clinical Psychiatry, Indiana University School of Medicine, and Adjunct Professor of Law, Indiana University McKinney School of Law, Indianapolis, IN. Address correspondence to: Philip Candilis, MD. E-mail: philip.candilis@dc.gov.

Disclosures of financial or other potential conflicts of interest: None.

the nature of peer review, hypothesis testing, and standards of examination. The corresponding array of legal cases, journals, texts, and organizational statements making up the standard of care can be a topic of inter- and intra-professional subjectivity and debate. Although standards are available for the organizational grading of clinical research (e.g., NIH, GRADE),^{7,8} forensic psychiatry has yet to pursue these approaches to assess the quality of its research outcomes. Consequently, the quality of forensic research in court remains deeply influenced by jurisdictional rules and culture, factfinder and adversary argumentation, and experts from diverse schools of thought. For example, psychiatrists, psychologists, and social workers may differ in their approaches to cases, as might psychodynamic and behaviorally oriented experts.

In 2020, the Research Committee of the American Academy of Psychiatry and the Law began discussions of a project that might address these limitations. Following consultation with the medical editor of the American Psychiatric Association's clinical practice guidelines, physician and epidemiologist Laura Fochtman, MD, MBI, the authors adopted the framework of a systematic review (i.e., a collection of empirical evidence using explicit, systematic methods) and the NIH grading criteria for a segment of forensic research that lent itself to quality assessment.

Original studies focusing on the process of adult inpatient competence restoration (i.e., restoration of competence to stand trial (RTC) or adjudicative competence) or among persons undergoing restoration were considered for the review. The topic was selected for three reasons. First, competence restoration is of ongoing relevance, as state mental health authorities and policymakers struggle with dramatically increasing referrals in recent years.⁹ Second, although there are many studies on RTC in the forensic literature, recent criticisms of competence restoration research have centered on the pervasiveness of descriptive research and single restoration groups rather than more rigorous comparisons of restored and unrestored persons.^{3,4} Third, the authors have conducted relevant empirical studies on the topic and can be considered content experts.¹⁰⁻¹³ By conducting this review, the authors intended to provide an overview of the current state of research on competence restoration to guide researchers, state mental health authorities, and policy makers on future examinations of the topic.

Methods

Two academic librarians qualified in systematic review conducted the initial search. Following guidance from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol,¹⁴ empirical studies published in English were identified in databases and reference lists from MEDLINE, PsycINFO, APA PsycArticles, Scopus, and the Cochrane Central Register of Controlled Trials. Search terms targeted adult inpatient competence restoration and its variations (online Appendix A). Doctoral dissertations were included, as were relevant references from primary articles identified by the search. If a dissertation was subsequently published, only the published version was selected for review.

After 244 studies were removed as duplicates, 1,187 studies from 1981 to 2022 were returned and uploaded to the Covidence platform ([covidence.org/reviews](https://www.covidence.org/reviews)), the screening and data extraction tool used by Cochrane authors.¹⁵ The titles and abstracts of all 1,187 studies were reviewed by each author in the initial round of screening, with the goal of removing publications that were clearly not related to competence restoration. Each author entered a vote on each study, as required by the Covidence platform, and conflicts were resolved by conference call. The authors then reviewed the abstracts of the remaining articles to ensure they were related to adult inpatient competence restoration, with each author again voting on each abstract and resolving difference by conference call. Many of the 1,187 articles described competence to make medical decisions and were consequently excluded, as were studies of juveniles. Such studies are identified in the PRISMA flow diagram (Fig. 1) as "wrong outcome" or "wrong population." Because the analysis centered on research articles that provided data on competence restoration, law review articles and thought and policy pieces were excluded. The full text of each remaining article was then reviewed independently by each author for relevance to adult inpatient competence restoration. There were 76 studies that met inclusion criteria (Fig. 1).

Each study type was categorized based on the NIH toolkit, which has rating forms for case series (the description of a group over time without a control group or randomization), case-control studies (a retrospective comparison of a study population with a similar control population), controlled interventions (comparison of groups receiving different interventions), pre-post studies (outcome analysis before

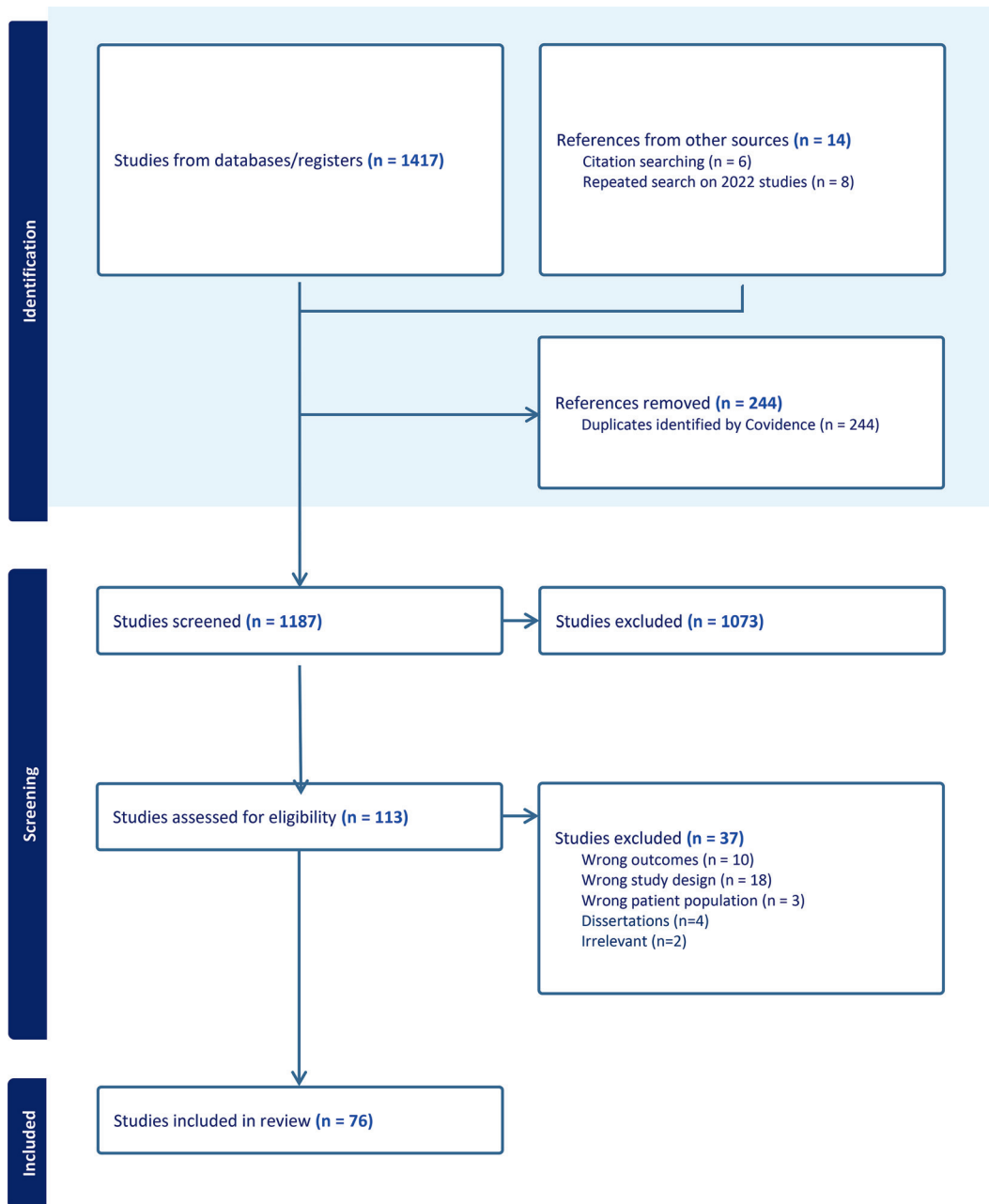


Figure 1. PRISMA flow diagram.

and after intervention), and meta-analysis/systematic reviews (data analysis from multiple studies on the same topic).

The NIH tools for tailored quality assessment were originally developed by the National Heart, Lung, and Blood Institute in 2013, focusing on the assessment of a study's internal validity.⁷ Tailored to specific study designs (e.g., case series, case-control, pre and post studies, meta-analyses; see Table 2), these tools prompt reviewers to determine whether a study has elements of design, analysis, and description

associated with lower risk of bias and higher likelihood of good quality. By prompting reviewers to judge the adequacy of sample descriptions and selection, sample and measure justification, and statistical adjustment of confounding variables, the tools provide a guide to discussions of the strength of a study's methods and are a mainstream approach to assessing the quality of human subject research. The reviewers must choose from Yes, No, or Other (cannot determine, not reported, not applicable) for each element. For all but the case series form, written guidance for answering

Quality of Competence Restoration Research

Table 1 General Descriptors for the Overall Quality Rating

	Design	Sample Size	Statistics
Good	Case control, meta-analysis, pre-post	>500	LCA, ROC, regression, power/effect analyses
Fair	Retrospective case series	100–500	Regression, correlation
Poor	Retrospective case series	<100	Descriptive or none, low or absent power

Note. Descriptors are the factors common to the completed NIH tools used for this analysis: e.g., “Good” studies used advanced statistics and case-control, meta-analytic, or pre-post designs.

Abbreviations: LCA = latent class analysis, ROC = receiver operating characteristic.

Table 2 Study Categories

Study Type	<i>N</i>
Case series	55
Case control	11
Meta-analysis	3
Survey	3
Controlled intervention	2
Pre-post	1
Other	1

each item is appended to each form. These tools were adopted as a resource for reviewers by The Journal in 2022.

At the end of each form, reviewers provided an overall quality rating of good, fair, or poor (Please see online Appendix B for a copy of the NIH tool for Case-Control Studies). The authors established a framework for the overall quality rating before reviewing individual studies. As described in Table 1, we agreed that study design, sample size, and statistical analysis were objective elements that could affect overall quality, alongside more subjective elements like the nature of the study question and whether the analysis was cogent. A case study could be considered good quality (and several were) if the sample size (*N*) were large and the statistical analysis were sophisticated; a case-control study could be poor quality if the *N* were small and the statistical analysis were limited.

After the 76 studies were identified and rated, the characteristics of each study were entered into a Microsoft Excel worksheet. These included the type of study, the *N*, the demographics of the study population (if available), the success of competence restoration (if available), the statistical methods, whether the authors initially agreed on the overall quality rating, and the final quality rating.

Results

Of the 76 studies qualifying for inclusion, 55 (72%) were identified as case series, with 11 (14%)

as case-control (Table 2). Eight (11%) of the 76 studies were classified as good quality, 54 (71%) as fair, and 14 (18%) as poor (Tables 3 and 4). Good studies were characterized by large samples, often in the thousands, and sophisticated statistical modeling like hierarchical regression and Latent Class Analysis, usually accompanied by power or effect analyses to justify their sample sizes (Table 4). Poor studies used small samples (often less than 20), few or no statistics (often because they were describing a new approach or offering commentary) and unclear sample selection and sample size justification.

Initial agreement between raters on the overall quality rating was 87 percent (66/76, kappa = .70). Fourteen studies (18%) used specific adjudicative competence tools: four used the Georgia Court Competency Tool (GCCT) while another four used the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA).¹⁶

Twenty-seven studies enrolled women (36%), while 35 studies enrolled majority non-White participants (46%, range 19 to 93%). Of 39 studies measuring restoration to competence, 32 (82%) reported restoration rates of over 50 percent (range 33 to 100%). A summary of study characteristics from all 76 studies can be found in online Appendix C.

Discussion

Implications for Research

A systematic review of the competence restoration literature showed that while there is extensive scholarship, only a small percentage of studies may be considered of good quality. Most studies, including those published by the authors, were of fair quality, had a modest *N*, and were retrospective in nature. Descriptive or simple correlational statistics predominated. Our findings are consistent with the conclusions of one high-quality study, Pirelli and Zapf’s 2020 attempted meta-analysis, which they famously abandoned because of the “dismal” quality of the

Table 3 Examples of Quality Categorization

	Studies	Description	Sample Size	Statistics
Good	Pirelli <i>et al.</i> , 2011 ¹⁷	Meta-analysis	>25,000	Inter-rater reliability, effect sizes, moderator effects
	Warren <i>et al.</i> , 2006 ¹⁸	Predictors of CR	8,416	Hierarchical logistic regression, effect sizes
Fair	Morris and Parker, 2008 ¹⁰	Retrospective review of factors affecting CR	1,475	Multiple logistic regression
	Secarea <i>et al.</i> , 2021 ¹²	Retrospective review of factors affecting CR	365	Multiple logistic regression
Poor	Ladds <i>et al.</i> , 1993 ¹⁹	Follow-up retrospective review of CR dispositions	68	Percentages only
	Herbel <i>et al.</i> , 2007 ²⁰	Novel assessment of duration of untreated psychosis in CR	22	None

Abbreviation: CR = competence restoration.

Table 4 Summary of Studies Rated “Good”

Lead Author	Year Published	N	Study Type	Strengths
Broderick ²¹	2020	20,040	Case series	Large database study of RTC defendants in one state. Used univariate analysis and then parametric regression to identify factors that affected LOS. Excellent description of statistical modeling.
Grossi ²²	2022	58	Case series	Creative and rigorous assessment of a new feigning instrument. Pilot study compared data from established tools to professional structured judgment using ANOVAs and hierarchical multiple regression.
Kivisto ²³	2020	492	Case series	Used LCA to identify the typology of a good-sized sample of defendants referred for RTC. An eight-class solution was identified based on 17 characteristics.
Pirelli ¹⁷	2011	26,139	Meta-analysis	Meta-analysis of 68 studies from 1967 to 2008. Discussion of publication bias, inter-rater reliability, effect sizes, and moderator effects.
Pirelli ⁴	2020	12,781	Meta-analysis	Attempted meta-analysis of 51 studies on RTC, 1975 to 2013. Did not complete analysis because of “dismal” quality of research. Provided a thorough description of the state of the literature and provided recommendations.
Warren ¹⁸	2006	8,416	Case series	Large study of CST evaluations from one state, 1993 to 2004. Studied multiple variables, determined effect sizes, and used chi-square and hierarchical logistic regression.
Warren ²⁴	2013	2,260	Case series	Large study of CST evaluations from one state, 1990 to 2007. Focused on reports which found ICST and offered an opinion about RTC. Used chi-square with Bonferroni correction and CHAID framework.
Williams ²⁵	1981	431	Case series	Historical article. Used difference in mean squares, discriminant analyses, and a novel weighted competency average, along with random sampling, reliability testing and cross-validation.

Abbreviations: ANOVA = analysis of variance, CHAID = chi squared automatic interaction detection, CST = competence to stand trial, ICST = incompetent to stand trial, LCA = latent class analysis, LOS = length of stay, RTC = restoration to competence.

research.⁴ It should be noted that Pirelli’s inaugural meta-analysis of research in this area, which merited a good quality rating in this analysis, was based on the many fair quality studies present in the literature at that time,¹⁷ so it is clear that there is value to conducting research of fair quality if that research can contribute to subsequent, more comprehensive research. Our systematic review using PRISMA, Covidence, and APA guidance differs from Pirelli’s statistical comparison of results and methods across studies (a meta-analysis), and offers a common alternative methodology for gathering empirical research.

The predominance of case series (i.e., the description of a group over time, without a control group or randomization) speaks to the ease of convenience samples: using one’s unit, program, or hospital to generate modest-sized samples within a familiar setting. Designing controlled studies with sufficient power to allow comparisons across jurisdictions remains aspirational, except in the analysis of large databases which usually require robust statistical and comparative methods. Even case-control studies that trace data backwards from a known outcome (and are therefore retrospective), may not be sufficient to

generate data that meaningfully identify the factors distinguishing restored from unrestored subjects. This point resonates with Pirelli and Zapf's call for pre-post designs, assessing competent and incompetent groups across time as they are restored (Ref. 4, p 154). Otherwise, subjects in restoration cannot technically be compared with restored subjects until the legal system makes its ultimate judgment.

The controlled design of noteworthy studies in this review serves as an indicator for the future of research in both adjudicative competence and forensic research in general (Tables 2 and 3). Researchers who wish to conduct research that will meet external quality criteria will need to consider two broader approaches: seeking out a large database or designing a prospective study with a significant *N*.

The former approach will likely be limited to researchers who have access to statewide databases. These databases exist in many states but are typically underutilized by the health authorities who collect the data. Developing a working relationship with the managers of these databases can lead to larger, more sophisticated investigations for both researchers and data owners within each state. Though comparisons across states would increase the data available for review, doing so would require matching or controlling for differences in statutes and procedures. Prospective research that applies power analysis (i.e., the calculation determining whether a sample is large enough to detect an effect if it exists) would be the goal of research that addresses current weaknesses in research methods.²⁶ Using a database or developing a controlled prospective study can overcome decades of data that have been more focused on the comparison of unrestored and competent groups rather than restored and unrestored.⁴

An almost universal component in the design of good studies, whether retrospective or prospective, is the contribution of expert statisticians who provide critical insight at every step of a research project, from conducting the power analysis to analyzing data with sophisticated modeling techniques. The descriptive and correlative statistics populating most adjudicative studies are insufficient for conclusions that draw on multiple variables and their interactions. This is the domain of multivariate regression employing more complex analytic methods.²⁷ Determining which factors have an effect while controlling for others advances research inquiry beyond the lesser capacities of description and correlation.

Although forensic psychiatrists typically have a limited background in statistics, those affiliated with academic departments and public sector agencies have access to professionals who work with these methods. Indeed, AAPL's Research Committee some years ago initiated a network of researchers who could support and consult with investigators seeking to expand their resources.

The absence of specific tools across studies is a related weakness for competence research. The most common tools in this review were the GCCT and MacCAT-CA.¹⁶ As research teams compare results across regions and jurisdictions, reliance on established, reliable tools becomes necessary for generalizing data. Though the GCCT is only a screening tool, its classic illustration of the courtroom offers recognizable cues to defendants and makes it popular across jurisdictions, while the MacCAT-CA relies on a voluminous decision-making literature supporting its exploration of a defendant's understanding, reasoning, and appreciation. Without implying a judgment on the preferability of one tool over another, the wider use of formal tools allows the comparison of methods across jurisdictions and the aggregation of data across comparable settings.

Unbalanced demographic representation is an associated finding of this study. With a minority of studies enrolling women, and the largest group of studies enrolling mostly non-White participants (46% of studies enrolled majority non-White participants), the research reflects the judicial system itself. Male-dominated samples generate results that do not necessarily generalize to women, whose diagnoses have different prevalence and symptom incidence.²⁸⁻³⁰ Women in forensic settings face unique challenges during pregnancy, lactation, and menopause alongside unique abusive, family, and economic influences.^{31,32} Research will have to generalize more broadly to achieve standards the National Institutes of Health set decades ago in its initiatives supporting women's health.³³

The predominance of non-White research samples reflects the U.S. forensic population itself, with detention rates, though decreasing, still highest among young Black and Hispanic men.³⁴⁻³⁷ Because study samples were generally small in this review, stratification by race and exploration of local variables were unavailable to dissect the influences of race, ethnicity, socioeconomic status, and neighborhood that are known to distinguish groups from one another.³⁸⁻⁴¹ Larger, representative samples can only

help overcome the literature's prevailing shortfalls of power and representation.

Implications for Training and Education

Forensic psychiatry has long focused on the process of conducting evaluations, preparing reports, and testifying in court, as these are the essence of forensic practice. All three elements revolve around the telling of stories: gathering the evaluatee's narrative during the interview and record review, placing the account in a written form that both describes the evaluatee and answers the questions posed by the attorney or judge, and then defending the account in testimony. These are the skills emphasized in forensic training as fellows build their professional identity. This emphasis is reflected in the research projects undertaken by forensic trainees and psychiatrists, many of which are about interesting evaluatees, essentially case studies or small case series. Annual meetings of the Academy in turn center on the cases and narratives that define the profession.

While case review remains a crucial element of the subspecialty, high quality research is equally important, for it is research that justifies how evaluations are conducted and which outcomes or interventions are recommended. The *Daubert* trilogy^{5,42,43} is a seminal part of the field, but forensic clinicians have little empirical data to offer to a *Daubert* challenge, even in competence restoration, which has been a focus of research for more than 50 years.

It would be salutary for training programs to focus their research efforts on robust research methods and projects, even though such projects may extend across multiple fellowship years. By doing so, a training program will advance the literature, teach cohorts of trainees, and underscore methods that can be applied broadly in future collaborations and meta-analyses. Indeed, the model of the systematic review, with which AAPL's Research Committee is familiar, offers a template for teams willing to examine other areas of empirical research. The committee itself can offer guidance.

The paucity of good research in one corner of the forensic literature can nonetheless be seen in a positive light: there are several worthy studies that set a standard of controlled designs, power analysis, large representative samples, and robust statistical methods. These should be attainable throughout the profession either by collaboration or targeted efforts,

contributing to the empirical foundations of the field and the community's confidence in expert opinions.

Acknowledgments

We are grateful for the expert guidance of Laura J. Fochtman, MD, MBI, and Paul Levett, the reference and instruction librarian at the Himmelfarb Health Sciences Library at the George Washington University School of Medicine. We extend our thanks as well to Elaine Sullo, former reference and instruction librarian of the Himmelfarb Library, and to Toni Yancey, the health sciences librarian of Saint Elizabeths Hospital in Washington, DC.

References

- 45 C.F.R. § 46.303(c) (2021)
- 45 C.F.R. § 46.306(a)(2) (2021)
- Heilbrun K, Giallella C, Wright HJ, *et al.* Treatment for restoration of competence to stand trial: Critical analysis and policy recommendations. *Psychol Pub Pol'y & L.* 2019; 25(4):266–83
- Pirelli G, Zapf PA. An attempted meta-analysis of the competency restoration research: Important findings for future directions. *Journal of Forensic Psychology Research and Practice.* 2020; 20(2):134–62
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993)
- Frye v. United States*, 293 F. 1013 (D.C. Cir 1923)
- US Dept. of Health and Human Services, National Institutes of Health. Assessing cardiovascular risk: Systematic evidence review from the risk assessment work group 2013. Available from: <https://www.nhlbi.nih.gov/sites/default/files/media/docs/risk-assessment.pdf>. Accessed November 14, 2023
- Guyatt GH, Oxman AD, Vist GE, GRADE Working Group, *et al.* GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008; 336(7650):924–6
- Callahan L, Pinals DA. Challenges to reforming the competence to stand trial and competence restoration system. *Psychiatr Serv.* 2020 Jul; 71(7):691–7
- Morris DR, Parker GF. Jackson's Indiana: State hospital competence restoration in Indiana. *J Am Acad Psychiatry Law.* 2008 Dec; 36(4):522–34
- Morris DR, Parker GF. Effects of advanced age and dementia on restoration of competence to stand trial. *Int'l J L & Psychiatry.* 2009 May-Jun; 32(3):156–60
- Secarea CM, Cleary SD, Candilis PJ. Factors affecting adjudicative competence and length of time to restoration. *J Forensic Sci.* 2021; 66(3):982–91
- Johnson NR, Candilis PJ. Outpatient competence restoration: A model and outcomes. *World J Psychiatry.* 2015; 5(2):228–33
- Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ.* 2021 Mar 29; 372:n71
- Covidence. How to work on a Cochrane review. Available from: <https://community.cochrane.org/help/tools-and-software/covidence>. Accessed February 6, 2023
- Zapf PA, Viljoen JL. Issues and considerations regarding the use of assessment instruments in the evaluation of competency to stand trial. *Behav Sci & L.* 2003; 21(3):351–6721
- Pirelli G, Gottdiener WH, Zapf PA. A meta-analytic review of competency to stand trial research. *Psychol Pub Pol'y & L.* 2011; 17(1):1–53

Quality of Competence Restoration Research

18. Warren JI, Murrie DC, Stejskal W, *et al.* Opinion formation in evaluating the adjudicative competence and restorability of criminal defendants: A review of 8,000 evaluations. *Behav Sci & L.* 2006; 24(2):113–32
19. Ladds B, Convit A, Zito J, Vitrai J. Involuntary medication of patients who are incompetent to stand trial: A descriptive study of the New York experience with judicial review. *Bull Am Acad Psychiatry Law.* 1993; 21(4):529–45
20. Herbel BL, Stelmach H. Involuntary medication treatment for competency restoration of 22 defendants with delusional disorder. *J Am Acad Psychiatry Law.* 2007 Mar; 35(1):47–59
21. Broderick C, Azizian A, Warburton K. Length of stay for inpatient incompetent to stand trial patients: Importance of clinical and demographic variables. *CNS Spectr.* 2020 Oct; 25(5):734–42
22. Grossi LM, Green D, Cabeldue M, Pivovarova E. Assessing feigning with the Feigning Evaluation Integrating Sources (FEINS) in a forensic psychiatric sample. *Psychol Serv.* 2022; 19(2):213–24
23. Kivisto AJ, Porter Staats ML, Connell R. Development and validation of a typology of criminal defendants admitted for inpatient competency restoration: A latent class analysis. *Law & Hum Behav.* 2020; 44(6):449–60
24. Warren JI, Chauhan P, Kois L, *et al.* Factors influencing 2,260 opinions of defendants' restorability to adjudicative competency. *Psychol Pub Pol'y & L.* 2013; 19(4):498–508
25. Williams W, Miller KS. The processing and disposition of incompetent mentally ill offenders. *Law & Hum Behav.* 1981; 5(4):245–61
26. Winter K, Pugh SL. An investigator's introduction to statistical considerations in clinical trials. *Urol Oncol.* 2019 May; 37(5):305–12
27. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis.* Upper Saddle River, NJ: Pearson Prentice Hall Pearson Education, Inc; 2019
28. Riecher-Rossler A. Sex and gender differences in mental disorders. *Lancet Psychiatry.* 2017; 4(1):8–9
29. McLean CP, Asnaani A, Litz BT, Hofmann SG. Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *J Psychiatr Res.* 2011; 45(8):1027–35
30. Angst J, Gamma A, Gastpar M. Depression Research in European Society Study, *et al.* Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. *Eur Arch Psychiatry Clin Neurosci.* 2002; 252(5):201–9
31. Streb J, Lutz M, Dudeck M, *et al.* Are women really different? Comparison of men and women in a sample of forensic psychiatric inpatients. *Front Psychiatry.* 2022 Mar 23; 13:857468
32. Friedman SH, Tamburello AC, Kaempf A, Hall RCW. Prescribing for women in corrections. *J Am Acad Psychiatry Law.* 2019 Dec; 47(4):476–85
33. The NIH Revitalization Act of 1993, Pub. L. No. 103-43 (1993)
34. Trestman RL. Is justice really blind? Nondominant groups in the American justice system. *J Am Acad Psychiatry Law.* 2018 Dec; 46(4):416–8
35. Pamplin JR, 2nd, Kelsall NC, Keyes KM, *et al.* Race, criminalization and urban mental health in the United States. *Curr Opin Psychiatry.* 2023 May 1; 36(3):219–36
36. Blankenship KM, Rosenberg A, Schlesinger P, *et al.* Structural racism, the social determination of health, and health inequities: The intersecting impacts of housing and mass incarceration. *Am J Public Health.* 2023 Jan; 113(S1):S58–S64
37. Robey JP, Massoglia M, Light MT. A generational shift: Race and the declining lifetime risk of imprisonment. *Demography.* 2023 Aug 1; 60(4):977–1003
38. Massey D. Residential segregation is the linchpin of racial stratification. *City Community.* 2016; 15(1):4–7
39. Nuru-Jeter AM, Michaels EK, Thomas MD, *et al.* Relative roles of race versus socioeconomic position in studies of health inequalities: A matter of interpretation. *Annu Rev Public Health.* 2018; 39:169–88
40. LaVeist TA. Disentangling race and socioeconomic status: A key to understanding health inequalities. *J Urban Health.* 2005 Jun; 82(2 Suppl 3):iii26–iii34
41. Candipan J, Phillips NE, Sampson RJ, Small M. From residence to movement: The nature of racial segregation in everyday urban mobility. *Urban Studies.* 2021; 58(15):3095–117
42. *General Electric Co. v. Joiner*, 522 U.S. 136 (1997)
43. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999)