

## REVIEW

# Punitive school discipline as a mechanism of structural marginalization with implications for health inequity: A systematic review of quantitative studies in the health and social sciences literature

Catherine dP. Duarte<sup>1,2</sup> | Candice Moses<sup>2</sup> | Melissa Brown<sup>3</sup> | Sandhya Kajeepeta<sup>4</sup> | Seth J. Prins<sup>4</sup> | Janelle Scott<sup>5</sup> | Mahasin S. Mujahid<sup>2</sup>

<sup>1</sup>Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, California, USA

<sup>2</sup>Division of Epidemiology, School of Public Health, University of California, Berkeley, California, USA

<sup>3</sup>Health and Social Behavior, School of Public Health, University of California, Berkeley, California, USA

<sup>4</sup>Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA

<sup>5</sup>Graduate School of Education, University of California, Berkeley, California, USA

## Correspondence

Catherine dP. Duarte, Department of Epidemiology & Population Health, Stanford University School of Medicine, 1701 Page Mill Rd, Stanford, CA 94305, USA.  
Email: [catherine\\_duarte@stanford.edu](mailto:catherine_duarte@stanford.edu)

## Funding information

Health Policy Research Scholars Program, Robert Wood Johnson Foundation

## Abstract

Punitive school discipline deploys surveillance, exclusion, and corporal punishment to deter or account for perceived student misbehavior. Yet, education and legal scholarship suggests it fails to achieve stated goals and exacerbates harm. Furthermore, it is disproportionately imposed upon Black, Latinx, Native/Indigenous, LGBTQIA, and disabled students, concentrating its harms among marginalized young people. Its implications for health, however, are less clear. Using public health theories of sociostructural embodiment, we propose a framework characterizing pathways linking societal ideologies (e.g., racism) to punitive discipline with implications for health and health inequity and then present our systematic review of the punitive school discipline–health literature ( $N = 19$  studies) conducted in accordance with PRISMA guidelines. Data were extracted on guiding theories, study characteristics, measurement, methods, and findings. This literature links punitive school discipline to greater risk for numerous health outcomes, including persistent depressive symptoms, depression, drug use disorder in adulthood, borderline personality disorder, antisocial behavior, death by suicide, injuries, trichomoniasis, pregnancy in adolescence, tobacco use, and smoking, with documented implications for racial health inequity. Using our adapted framework, we contextualize results and recommend avenues for future research. Our findings support demands to move away from punitive school discipline toward health-affirming interventions to promote school connectedness, safety, and wellbeing.

## KEYWORDS

health, health inequity, punishment, racism, school discipline, structural marginalization, suspension

## INTRODUCTION

Education is a well-established social determinant of health. Yet, much of this literature comprises studies that link higher attainment to better health.<sup>1,2</sup> Fewer studies have examined characteristics of the educational environment that may influence this association,<sup>3</sup> though these studies consistently challenge conceptualizations of education as universally health-benefiting and suggest school spaces themselves can be sites of harm.<sup>3,4</sup> One such characteristic shaping educational experiences—with plausible links to health outcomes—is punitive school discipline.

Punitive school discipline is a process that deploys surveillance (e.g., metal detectors and armed campus officers), restrictive (e.g., restraint and seclusion), physical (e.g., corporal punishment), and exclusionary (e.g., removal from school) tactics to account for or deter perceived student misbehavior.<sup>5</sup> Most often, this is operationalized via retributive mechanisms that remove students from their classrooms or schools,<sup>5</sup> including suspensions, expulsions, and school-based law enforcement referrals/arrests.<sup>6</sup> While purported to increase or maintain school safety, evidence suggests punitive approaches not only fail to achieve this end (i.e., suspensions do not meet the stated target outcome of deterring future perceived misbehavior),<sup>7</sup> but may actually exacerbate harm.<sup>7–13</sup>

Still, exposure to punitive discipline is common for young people attending schools in the United States. From kindergarten through high school, over one-third of all public school students experience at least one out-of-school suspension;<sup>14</sup> 19 states still permit corporal punishment in schools;<sup>15</sup> and over half of public schools report having a campus-based armed law enforcement officer<sup>16</sup> such that over 10 million students attend schools with police, but no counselor, nurse, psychologist, or social worker.<sup>17</sup> Notably, a substantial body of literature finds that students who are structurally marginalized—specifically, Black, Latinx, Native/Indigenous, disabled young people, and lesbian, gay, bisexual, transgender, queer, intersex, and asexual (LGBTQIA) individuals—are at increased risk of exposure to punitive discipline and school-based policing, beginning as early as preschool and persisting across grades and school settings, indicating this is a pathway through which discrimination operates in young people's lives.<sup>7,9,18–22</sup> While these data speak to the expansive reach and stark inequities in the deployment of punitive discipline, they likely underestimate its true scale given many forms of punitive school discipline are not systematically documented (e.g., restraint<sup>23</sup> and “soft expulsion”<sup>24</sup>).

Punishment, however, has not always characterized discipline in US schools. Rather, through a series of education policy mandates at federal, state, and local levels, there was an orchestrated investment in punitive school discipline—independent of any changes in student behavior—beginning in the late 1980s and largely reinforced through the early 2000s.<sup>9,11,19</sup> For example, by 1997, 94% of public schools had adopted “zero tolerance” policies—whereby particular types of perceived student misbehavior (e.g., related to drug use, weapons, and school disruption) had mandated punitive consequences that neither considered context nor examined potential causes.<sup>25</sup> This shift mirrored other increasingly punitive legal policies and practices

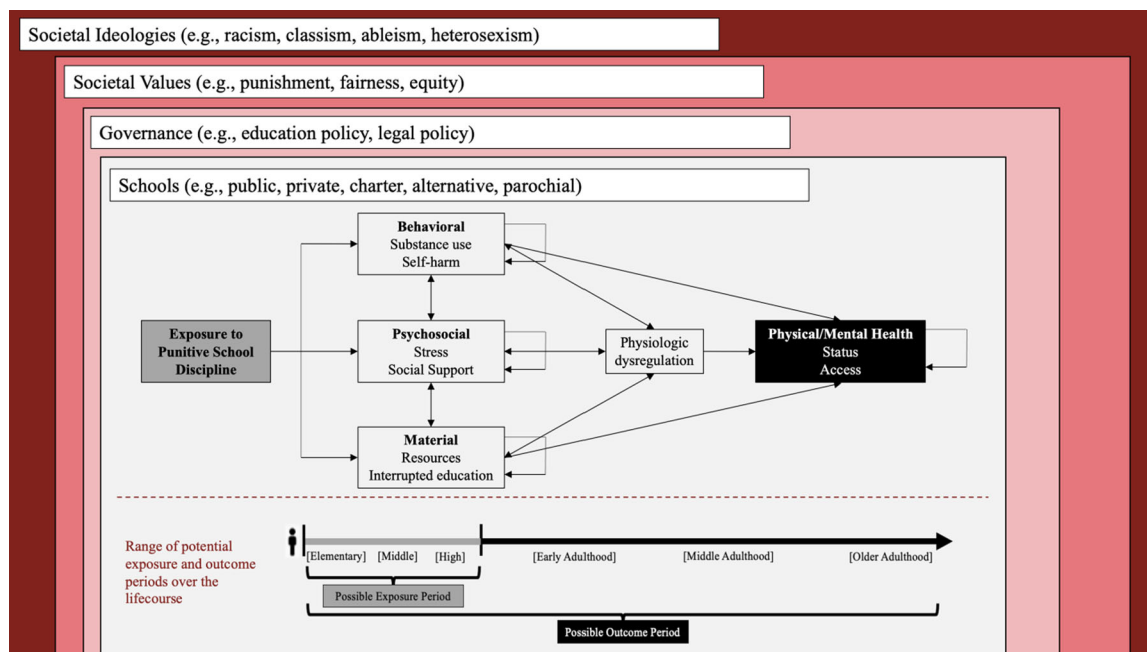
implemented at this time (e.g., Broken Windows Policing<sup>7</sup> and War on Drugs<sup>9,19</sup>), indicating a mobilization across institutions of governance toward punitive, retributive, and exclusionary tactics. As geographer Ruth Wilson Gilmore notes, these social policies were responses to the “organized abandonment” neoliberal transformation of the state.<sup>26</sup> The state withdrew from guaranteeing material security, and as criminologist Jonathan Simon describes, began to “govern through crime” (i.e., manage social consequences of that withdrawal through punishment across public institutions, such as the welfare system and public education) with a disproportionate targeting of marginalized people and communities.<sup>27</sup>

In response, affected communities began organizing to document the implications and intervene on the expansion of punitive practices in schools, prompting studies among education and legal scholars. Those studies found consistent associations between punitive school discipline and outcomes like reductions in student academic performance, lower educational attainment, and pushout/pullout into the criminal legal system.<sup>7,28</sup> However, the extent to which school discipline's impact on the health of US students has been studied, and what those studies collectively conclude, is unclear. Especially given its prevalence, documenting the health implications of punitive school discipline is critical to (1) identifying potential causes of population health and health inequity; (2) informing ongoing policy and practice discussions; and (3) mobilizing the fields of public health and clinical medicine to become involved with those deliberations, addressing health in all policies and practices.<sup>29,30</sup>

## Conceptual framework: Punitive discipline as a pathway through which structural determinants shape health

Drawing on existing frameworks and theories of sociostructural embodiment,<sup>31–38</sup> we propose a conceptual framework characterizing hypothesized pathways linking societal ideologies to punitive school discipline with implications for health and health inequity, both immediately as well as into adulthood (Figure 1).

Specifically, *Societal Ideologies* (e.g., racism) prescribe and normalize *Societal Values* around punishment as a means of accountability or social control (i.e., stigmatizing difference, assigning labels of “deviance,” and seeking to forcibly impose conformity).<sup>39–42</sup> This then guides the materialization of education policy (*Governance*, Figure 1) that is implicitly or explicitly designed to criminalize structurally marginalized students (e.g., Black and Brown students, poor students, disabled students, and LGBTQIA students).<sup>39,41,43</sup> Within school environments (*Schools*), the implementation of these policies—compounded by individual-level bias—results in structurally marginalized students being at increased risk for punitive school discipline, which is then justified by attributing perceived student misbehavior to an inherent characteristic of the student (e.g., “troublemaker”), their peers (e.g., “bad crowd”), or their communities (e.g., “high violence”).<sup>5,42</sup> This structure (*Societal Ideologies*, *Societal Values*, *Governance*, and *Schools*) envelopes material, psychosocial, and



**FIGURE 1** Punitive discipline as a pathway through which structural determinants shape health: an adapted conceptual framework. Adapted/informed by Refs. 31–38.

behavioral pathways, the implications of which are two-fold: first, that this structure exerts direct influence over these pathways; second, that the health effects of each pathway are compounded by the *Societal Ideologies* and *Societal Values* manifesting in concert across systems of *Governance* (e.g., legal policy, health policy, and housing policy).<sup>43,44</sup> Finally, as with the frameworks on which it draws, the nested structure of this framework acknowledges the interdependence of its layers. For example, just as ideologies shape governance, governance structures and schools reify ideology. Similarly, just as the design of governance structures marginalize populations, so too do individuals organize and movement build to resist and shift those structures.<sup>31</sup>

Given that schools are foundational environments in young people's lives and identity formation—from the overall time spent in schools, to peer/teacher relationships nourished there, and beyond<sup>45</sup>—exclusionary practices that disrupt access to these environments may have implications for lifecourse health. Material pathways may include the disruption of education trajectories, with lower educational attainment linked to poorer health experiences through, for example, reductions in income or greater barriers to accessing healthcare.<sup>1,2,46–49</sup> Psychosocial pathways may result from punitive discipline triggering repeated stress reactions that can lead to physiologic dysregulation, premature biological aging, and/or adverse mental health outcomes.<sup>36,38,50,51</sup> Behavioral pathways may include health-harming practices, like substance misuse or self-harm as coping responses to punitive discipline and associated stigma.<sup>52,53</sup> They may also result from punitive school exclusion leading to increases in recreational substance use or perceived truancy (both considered “status offenses”), with implications for health and wellbeing either directly, or via increased legal system targeting.<sup>54</sup> The adapted con-

ceptual framework reflects that these pathways likely operate over different periods throughout the lifecourse, with health implications presenting from childhood through older adulthood (i.e., immediate, persistent, or latent onset).<sup>2,38,55,56</sup> Further, exposure during developmentally sensitive periods could exacerbate health effects.<sup>2,38</sup> Taken together, this adapted framework suggests that systems of structural marginalization (e.g., racism) shape differential exposure to punitive school discipline in early life resulting in inequitable risk for a variety of adverse health outcomes over the lifecourse.

## Objective

While individual studies have examined the association between punitive discipline and health, we identified only one review of that literature.<sup>57</sup> However, that review (published in 2014) focused on a specific health outcome (psychiatric disorder or impairing psychopathology) and summarized an international literature. Given the implications of each country's unique history of structural marginalization—and consequently, differences in the nature of punitive policy, the overall frequency of school-based practice, and differential targeting of students—this could obscure US-specific phenomena. Other related reviews have considered broader constructs, like the “school environment” though these have not explicitly examined punitive school discipline.<sup>3</sup>

We present a systematic review of the health and social sciences literature on the relationship between punitive school discipline and health among US-based school-aged young people. Specifically, we define the scope of research on punitive school discipline and

health, how indicators of discipline and health have been conceptualized and operationalized, if an association has been documented, and to what extent the role of structural marginalization has been examined.

## MATERIALS AND METHODS

Our search strategy was developed in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.<sup>58</sup>

### Information sources

We conducted a systematic review of primary electronic databases indexing public health, education, criminology, and social sciences literature; specifically, PubMed, Sociological Abstracts, ERIC, and PsycINFO. Our initial search was conducted on August 17, 2019, and a second search was conducted on July 21, 2020.

### Search strategy

With a University-based reference librarian, we developed a search string to identify publications focused on (1) school discipline and (2) health among (3) US-based, (4) school-aged (5) young people. Where appropriate, search terms were separated by Boolean operators and truncated using an asterisk (e.g., *adolescen\** to include adolescence, adolescent, and adolescents). The first author, C.dP.D., conducted the database searches. For a detailed description of the search string, see [Supplementary Materials](#).

### Additional records through other sources

An independent search, screening titles and abstracts, was concurrently performed by a collaborating author, S.K., in PubMed, Education Research Complete, ERIC, and PsychINFO between September 9, 2019 and October 2, 2019.

### Eligibility criteria

We began with a broader set of inclusion criteria: abstract available; English-language publication; US-based; primary school discipline indicator; primary health status, behavior, access, or policy indicator; school discipline occurred in a prekindergarten through the high-school-aged population; and involved original, peer-reviewed, empirical research (Stage 1). To streamline methodological approaches, we then narrowed the eligibility criteria such that qualitative studies were excluded (Stage 2). We then further restricted to the particular

literature of interest, excluding analyses which were: descriptive only; focused on discipline as an outcome; or focused only on distal health indicators (e.g., “pleased with a partner’s pregnancy”) (Stage 3).

### Selection process

We uploaded the search yield to the Covidence systematic review software,<sup>59</sup> manually confirming the software’s automatic identification and removal of duplicates. In Stage 1, we first piloted the eligibility screening tool we developed, then at least two reviewers used the finalized tool to assess each abstract. Abstracts with affirmative concordance or affirmative discordance (“Yes,” “Maybe”) were advanced to Stage 2. Abstracts with negative concordance (“No”) were excluded. Abstracts with conflicting votes were discussed by all reviewers until reaching a consensus. Abstracts with insufficient information to make a determination were advanced to Stage 2. In Stage 2, full-text publications were rescreened using the narrowed screening tool. In Stage 3, full-text publications were again rescreened using the further restricted screening tool.

### Data collection

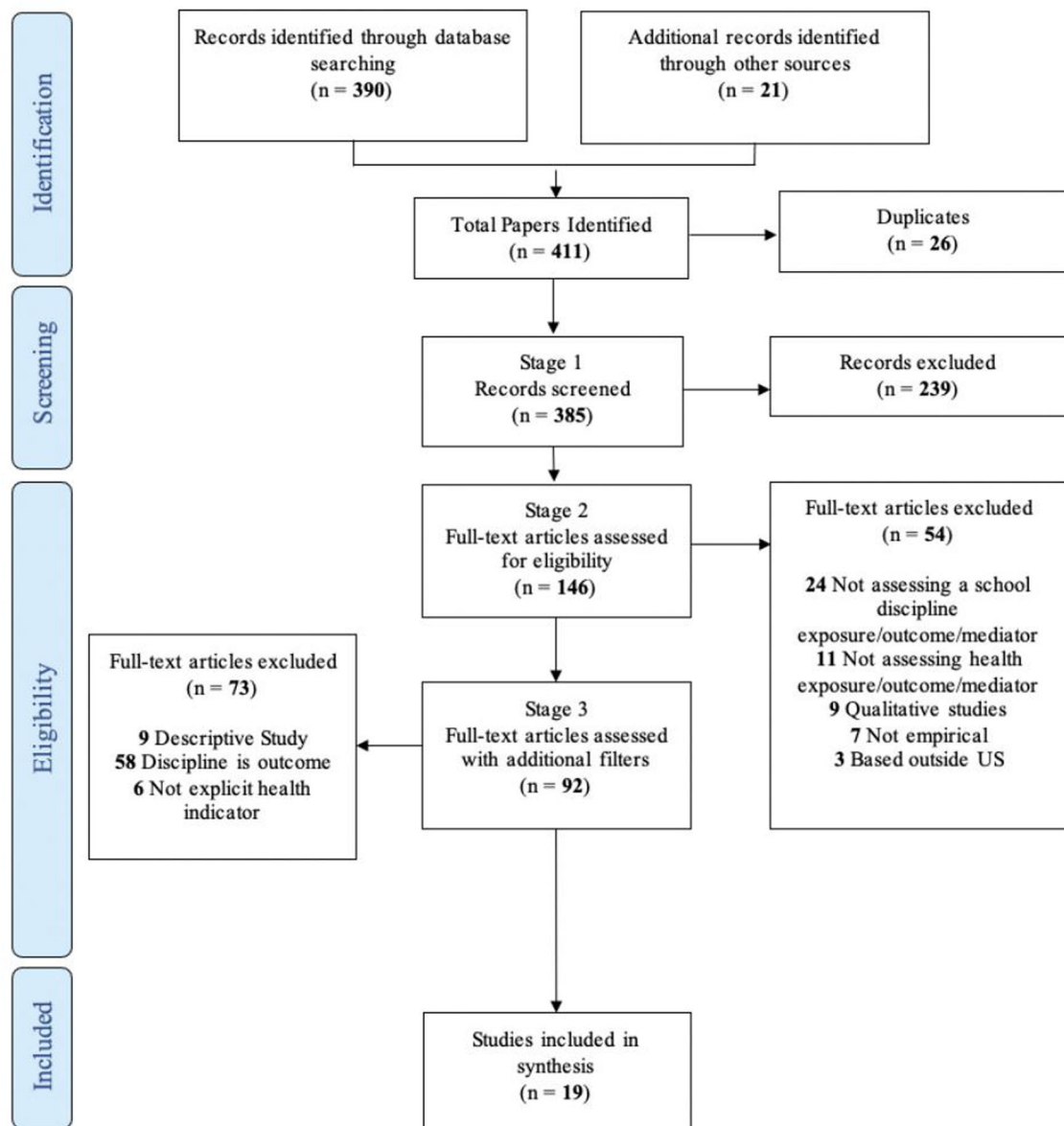
We developed and piloted a template for data extraction and then programmed the template to Covidence. Each publication was extracted by at least two reviewers (the first author, C.dP.D., served as a reviewer on all publications). Reviewers met to discuss discordance and confirm final extractions.

### Data items

We extracted data on: (1) publication information (study title, authors, publication year, and funding source); (2) theory (whether the guiding theory was explicitly identified); (3) study characteristics (data source, analytic sample size, sample description, and study period); (4) measurement (operationalization/conceptualization of school discipline and health); (5) methodology (study design, length of follow-up, level of analysis, covariates, effect modifiers, and study limitations); and (6) findings (magnitude and precision). If any information was missing or unclear, efforts were made to identify or confirm (e.g., using descriptions of data sources to identify secondary datasets).

### Synthesis methods

The first author, C.dP.D., summarized key characteristics across each publication and then conducted a thematic synthesis to identify patterns, and from those patterns document prevailing themes that emerged across this literature.



**FIGURE 2** PRISMA flow diagram.

## RESULTS

### Study characteristics

We screened 411 titles and abstracts, with 26 excluded as duplicates and 239 excluded in Stage 1. We then screened the remaining 146 full-text manuscripts, with 54 excluded in Stage 2 and 73 removed in Stage 3. Figure 2 provides a detailed summary of this multistage review process. Ultimately, 19 manuscripts comprised the final yield. In Table 1, we summarize the distribution of key characteristics across all studies. Tables 2 and 3 present these characteristics by individual study and Table 4 describes the operationalization of each discipline and health indicator.

### Characteristics of theoretical frames

To examine assumptions underlying the authors' analysis and/or interpretation of the school discipline–health association, we extracted data on whether a guiding theory was explicitly identified. Nearly half of publications did so, with notable variability in which theories were named and some studies invoking more than one (Table 1). We examined whether theories described mechanisms that were structural (i.e., invoking systems), contextual (i.e., invoking interpersonal interactions that shape immediate environments), or inherent (i.e., invoking biologically or psychologically essentialist attributes). We found that most studies posited contextual and/or inherent mechanisms (87.5%). The individual and interpersonal processes that are defining features

**TABLE 1** Summary of included studies,  $N=19$ 

Study characteristic	N	%	References
<b>Theory</b>			
Explicit guiding theory	8	42.1%	Cottrell <sup>75</sup> ; Fothergill et al. <sup>67</sup> ; Gillmore et al. <sup>63</sup> ; Hemphill et al. <sup>79</sup> ; Lee et al. <sup>65</sup> ; Paek et al. <sup>66</sup> ; Racz et al. <sup>60</sup> ; Rosenbaum <sup>69</sup>
Structural mechanisms	1	12.5%	Paek et al. <sup>66</sup>
Contextual mechanisms	6	75.0%	Cottrell <sup>75</sup> ; Fothergill et al. <sup>67</sup> ; Gillmore et al. <sup>63</sup> ; Hemphill et al. <sup>79</sup> ; Lee et al. <sup>65</sup> ; Rosenbaum <sup>69</sup>
Inherent mechanisms	3	37.5%	Gillmore et al. <sup>63</sup> ; Lee et al. <sup>65</sup> ; Racz et al. <sup>60</sup>
<b>Data source</b>			
Primary	4	21.1%	Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Kramer et al. <sup>76</sup> ; Racz et al. <sup>60</sup>
Secondary	14	73.7%	Alexander et al. <sup>80</sup> ; Bottiani et al. <sup>64</sup> ; Cottrell <sup>75</sup> ; Elster et al. <sup>62</sup> ; Fothergill et al. <sup>67</sup> ; Fowler et al. <sup>81</sup> ; Gillmore et al. <sup>63</sup> ; Lee et al. <sup>65</sup> ; Paek et al. <sup>66</sup> ; Ramey <sup>61</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup> ; Slade <sup>82</sup> ; Talluri et al. <sup>78</sup>
Both	1	5.3%	Gould et al. <sup>68</sup>
<b>Sample size</b>			
Min	164		Kramer et al. <sup>76</sup>
Max	19,726		Bottiani et al. <sup>64</sup>
Median	1248		–
Mean (SD)	3346.6 (4807.0)		–
<b>Location</b>			
National	7	36.8%	Cottrell <sup>75</sup> ; Elster et al. <sup>62</sup> ; Fowler et al. <sup>81</sup> ; Ramey <sup>61</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup> ; Slade <sup>82</sup>
State	6	31.6%	Bottiani et al. <sup>64</sup> ; Gould et al. <sup>68</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Lee et al. <sup>65</sup> ; Paek et al. <sup>66</sup>
Local	6	31.6%	Alexander et al. <sup>80</sup> ; Fothergill et al. <sup>67</sup> ; Gillmore et al. <sup>63</sup> ; Kramer et al. <sup>76</sup> ; Racz et al. <sup>60</sup> ; Talluri et al. <sup>78</sup>
<b>Age composition</b>			
Middle school-aged	3	15.8%	Alexander et al. <sup>80</sup> ; Fothergill et al. <sup>a 67</sup> ; Talluri et al. <sup>78</sup>
High school-aged	5	26.3%	Bottiani et al. <sup>64</sup> ; Elster et al. <sup>62</sup> ; Lee et al. <sup>65</sup> ; Paek et al. <sup>66</sup> ; Racz et al. <sup>60</sup>
Elementary and middle school-aged	2	10.5%	Ramey <sup>61</sup> ; Slade <sup>82</sup>
Middle and high school-aged	7	36.8%	Cottrell <sup>75</sup> ; Gillmore et al. <sup>63</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Kramer et al. <sup>76</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup>
Elementary through high school-aged	2	10.5%	Fowler et al. <sup>81</sup> ; Gould et al. <sup>68</sup>
<b>Gender composition<sup>b</sup></b>			
Boys only	2	10.5%	Racz et al. <sup>60</sup> ; Ramey <sup>61</sup>
Girls only	2	10.5%	Elster et al. <sup>62</sup> ; Gillmore et al. <sup>63</sup>
Boys and girls	15	78.9%	Alexander et al. <sup>80</sup> ; Bottiani et al. <sup>64</sup> ; Cottrell <sup>75</sup> ; Fothergill et al. <sup>67</sup> ; Fowler et al. <sup>81</sup> ; Gould et al. <sup>68</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Kramer et al. <sup>76</sup> ; Lee et al. <sup>65</sup> ; Paek et al. <sup>66</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup> ; Slade <sup>82</sup> ; Talluri et al. <sup>78</sup>
<b>Racial composition</b>			
Black only	1	5.3%	Fothergill et al. <sup>67</sup>
Latinx only	1	5.3%	Talluri et al. <sup>78</sup>
Black and White	4	21.1%	Alexander et al. <sup>80</sup> ; Bottiani et al. <sup>64</sup> ; Cottrell <sup>75</sup> ; Paek et al. <sup>66</sup>
Black, White, and Latinx	7	36.8%	Elster et al. <sup>62</sup> ; Fowler et al. <sup>81</sup> ; Gould et al. <sup>68</sup> ; Racz et al. <sup>60</sup> ; Ramey <sup>61</sup> ; Rosenbaum <sup>69</sup> ; Slade <sup>82</sup>
White and “Non-White”	2	10.5%	Gillmore et al. <sup>63</sup> ; Kramer et al. <sup>76</sup>
Did not restrict by race	3	15.8%	Hemphill et al. <sup>79</sup> ; Lee et al. <sup>65</sup> ; Rushton et al. <sup>74</sup>
Not specified	1	5.3%	Hemphill et al. <sup>77</sup>

(Continues)

**TABLE 1** (Continued)

Study characteristic	N	%	References
<b>Study design</b>			
Cross-sectional	6	31.6%	Alexander et al. <sup>80</sup> ; Cottrell <sup>75</sup> ; Fowler et al. <sup>81</sup> ; Gillmore et al. <sup>63</sup> ; Lee et al. <sup>65</sup> ; Racz et al. <sup>60</sup>
Cohort	11	57.9%	Bottiani et al. <sup>64</sup> ; Elster et al. <sup>62</sup> ; Fothergill et al. <sup>67</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Paek et al. <sup>66</sup> ; Ramey <sup>61</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup> ; Slade <sup>82</sup> ; Talluri et al. <sup>78</sup>
Case-control	2	10.5%	Gould et al. <sup>68</sup> ; Kramer et al. <sup>76</sup>
<b>Study period</b>			
Min (years)	1		Alexander et al. <sup>80</sup> ; Cottrell <sup>75</sup> ; Fowler et al. <sup>81</sup> ; Gillmore et al. <sup>63</sup> ; Kramer et al. <sup>76</sup> ; Lee et al. <sup>65</sup> ; Racz et al. <sup>60</sup>
Max (years)	37		Fothergill et al. <sup>67</sup>
Median (years)	2		Bottiani et al. <sup>64</sup> ; Gould et al. <sup>68</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Paek et al. <sup>66</sup> ; Rushton et al. <sup>74</sup> ; Slade <sup>82</sup>
Mean (SD)	5.7 (10.6)		–
<b>Categorical</b>			
Turn toward mandatory punishment (~Up to 1990)	5	26.3%	Alexander et al. <sup>80</sup> ; Elster et al. <sup>62</sup> ; Fowler et al. <sup>81</sup> ; Gillmore et al. <sup>63</sup> ; Gould et al. <sup>68</sup>
Height of mandatory punishment (~1991–2011)	10	52.6%	Cottrell <sup>75</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Kramer et al. <sup>76</sup> ; Paek et al. <sup>66</sup> ; Racz et al. <sup>60</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup> ; Slade <sup>82</sup> ; Talluri et al. <sup>78</sup>
Considering alternatives to punishment (~2012–present)	2	10.5%	Bottiani et al. <sup>64</sup> ; Lee et al. <sup>65</sup>
Multiperiod	2	10.5%	Fothergill et al. <sup>67</sup> ; Ramey <sup>61</sup>
<b>Follow-up</b>			
Single time point	7	36.8%	Alexander et al. <sup>80</sup> ; Cottrell <sup>75</sup> ; Fowler et al. <sup>81</sup> ; Gillmore et al. <sup>63</sup> ; Lee et al. <sup>65</sup> ; Racz et al. <sup>60</sup> ; Slade <sup>82</sup>
Prospective	11	57.9%	Bottiani et al. <sup>64</sup> ; Elster et al. <sup>62</sup> ; Fothergill et al. <sup>67</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Paek et al. <sup>66</sup> ; Ramey <sup>61</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup> ; Slade <sup>82</sup> ; Talluri et al. <sup>78</sup>
Retrospective	2	10.5%	Gould et al. <sup>68</sup> ; Kramer et al. <sup>76</sup>
<b>Level of analysis</b>			
Multilevel	3	15.8%	Bottiani et al. <sup>64</sup> ; Lee et al. <sup>65</sup> ; Paek et al. <sup>66</sup>
Individual	16	84.2%	Alexander et al. <sup>80</sup> ; Cottrell <sup>75</sup> ; Elster et al. <sup>62</sup> ; Fothergill et al. <sup>67</sup> ; Fowler et al. <sup>81</sup> ; Gillmore et al. <sup>63</sup> ; Gould et al. <sup>68</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Kramer et al. <sup>76</sup> ; Racz et al. <sup>60</sup> ; Ramey <sup>61</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup> ; Slade <sup>82</sup> ; Talluri et al. <sup>78</sup>
<b>Discipline indicator</b>			
<b>Individual-level</b>			
Suspensions	8	42.1%	Cottrell <sup>75</sup> ; Elster et al. <sup>62</sup> ; Fothergill et al. <sup>67</sup> ; Gould et al. <sup>68</sup> ; Hemphill et al. <sup>77</sup> ; Hemphill et al. <sup>79</sup> ; Rosenbaum <sup>69</sup> ; Rushton et al. <sup>74</sup>
Expulsions	1	5.3%	Racz et al. <sup>60</sup>
Suspensions/expulsions	5	26.3%	Fowler et al. <sup>81</sup> ; Gillmore et al. <sup>63</sup> ; Kramer et al. <sup>76</sup> ; Ramey <sup>61</sup> ; Slade <sup>82</sup>
Detention/suspension	1	5.3%	Talluri et al. <sup>78</sup>
Sent to principal's office	1	5.3%	Alexander et al. <sup>80</sup>
<b>School-level</b>			
Black-White discipline gap	1	5.3%	Bottiani et al. <sup>64</sup>
Suspension rate	1	5.3%	Lee et al. <sup>65</sup>
Policy enforcement	1	5.3%	Paek et al. <sup>66</sup>

(Continues)



**TABLE 1** (Continued)

Study characteristic	N	%	References
<b>Health indicator</b>			
Mental health	6	31.6%	<i>Mental Health Indicators:</i> Cottrell <sup>75</sup> ; Kramer et al. <sup>76</sup> ; Rushton et al. <sup>74</sup> <i>Critiqued Mental Health Indicators:</i> Bottiani et al. <sup>64</sup> ; Hemphill et al. <sup>77</sup> ; Lee et al. <sup>65</sup>
Behavioral health	6	31.6%	Fothergill et al. <sup>67</sup> ; Gillmore et al. <sup>63</sup> ; Hemphill et al. <sup>79</sup> ; Paek et al. <sup>66</sup> ; Racz et al. <sup>60</sup> ; Talluri et al. <sup>78</sup>
Physical health	5	26.3%	Alexander et al. <sup>80</sup> ; Elster et al. <sup>62</sup> ; Fowler et al. <sup>81</sup> ; Gould et al. <sup>68</sup> ; Rosenbaum <sup>69</sup>
Health system involvement	2	10.5%	Ramey <sup>61</sup> ; Slade <sup>82</sup>
<b>Effect modification</b>			
Gender	3	15.8%	<i>Stratified:</i> Alexander et al. <sup>80</sup> <i>Interaction:</i> Fothergill et al. <sup>67</sup> ; Gould et al. <sup>68</sup>
Race/ethnicity	5	26.3%	<i>Stratified:</i> Bottiani et al. <sup>64</sup> ; Cottrell <sup>75</sup> ; Ramey <sup>61</sup> ; Slade <sup>82</sup> <i>Interaction:</i> Gould et al. <sup>68</sup>
Age	1	5.3%	<i>Interaction:</i> Gould et al. <sup>68</sup>
Grade	1	5.3%	<i>Stratified:</i> Hemphill et al. <sup>79</sup>
Socioeconomic status	2	10.5%	<i>Stratified:</i> Fowler et al. <sup>81</sup> <i>Interaction:</i> Gould et al. <sup>68</sup>
Geography	2	10.5%	<i>Stratified:</i> Elster et al. <sup>62</sup> <i>Interaction:</i> Hemphill et al. <sup>79</sup>

<sup>a</sup>Participants were in grade 1 at recruitment and grade 7/8 when exposure was ascertained.

<sup>b</sup>All studies examined gender as a binary; some studies used sex and gender interchangeably.

of these contextual and inherent mechanisms were also reflected in the authors' analytic approaches: most being at the individual-level, forgoing structural analyses. Insofar as analyses examining variation across experiences of structural marginalization (e.g., by race and gender) cited theory, they exclusively focused on inherent or contextual processes. None invoked theories with an explicit examination of, for example, structural racism or intersectional axes of structural oppression.

## Datasets and sample descriptions

Studies focused on a range of US geographies, including national (36.8%), state (31.6%), and localized levels (i.e., regions, cities, and institutions; 31.6%) (Table 1). Most analyses (73.7%) were conducted in secondary datasets. Accounting for within-study exclusions, analytic sample sizes ranged from 164 to 19,726 participants (median: 1248; mean: 3346.6; and SD: 4807.0). Most studies were restricted to participants who identified as Black, White, or Latinx (36.8%) or to participants who identified as Black or White (21.1%). The remaining studies: did not restrict by race/ethnicity (15.8%); compared White to "non-White" participants (10.5%); were racially homogenous (5.3% with Black participants only and 5.3% with Latinx participants only); or did not specify sample racial demographics (5.3%). Two studies were restricted to boys,<sup>60,61</sup> while two others were restricted to girls.<sup>62,63</sup> For most studies, exposure was assessed among a mix of middle and high school students (36.8%) or high school students alone (26.3%).

The mean study period length was 5.7 years (median: 2 years; range: 1–37 years of follow-up).

## Methodology

More than half of the studies used a prospective cohort design (57.9%), 31.6% were cross-sectional, and 10.5% were retrospective case-control studies (Table 1). For most studies (84.2%), exposure and outcome were measured and analyzed at the individual-level, with the remaining (15.8%) examining group-level exposures and individual outcomes.

## Measurement

"School discipline" was primarily operationalized as experiences of suspension and/or expulsion (Table 1). Comparison groups were largely comprised of students with no measured experience of punitive discipline. Only three studies examined school-level measures (Black-White discipline gap;<sup>64</sup> suspension rate;<sup>65</sup> and policy enforcement practices<sup>66</sup>). In most studies, exposure was measured by student self-report (68.4%), with fewer using parental/caregiver report (26.3%), secondary administrative data (10.5%), or school administrator report (10.5%). Two studies used multiple sources for exposure assessment,<sup>67,68</sup> so percentages do not sum to 100%.



**TABLE 2** Study characteristics of individual included studies, N=19

Reference	Explicit guiding theory	Data source	Size (N)	Location	Population	Study period	Study design	Follow-up	Level of analysis
Alexander et al. (1995)	–	Secondary	758	Rural eastern shore, Maryland	Black and White public school students in grade 8	1987	Cross-sectional	Single time point	Individual
Bottiani et al. (2017)	–	Secondary	19,726	Maryland (Suburban/Rural)	Black and White public high school students, mean aged 15.9 years	2011–2013	Cohort	Prospective	Multilevel
Cottrell (2018)	General strain theory	Secondary	3289	National	Black and White middle and high school adolescents	1994–1995	Cross-sectional	Single time point	Individual
Elster et al. (1990)	–	Secondary	1651	National	Black, White, and Hispanic young women aged 15–17 years	1980–1985	Cohort	Prospective	Individual
Fothergill et al. (2008)	Life course social field theory; Strain theory; Social control theory; Primary socialization theory	Secondary	1053	Chicago, Illinois	African American public and parochial school students followed since grade 1 (age 6–42 years) <i>Note. Participants were in grade 7 or 8 when exposure was ascertained</i>	1966–2003	Cohort	Prospective	Individual
Fowler et al. (1992)	–	Secondary	10,362	National	Black, White, and Hispanic children and adolescents in grades 1 through 12 aged 5–17	1988	Cross-sectional	Single time point	Individual
Gillmore (1992)	Problem behavior theory; Social development model	Secondary	241	Large metropolitan area, Northwest US	White, non-Hispanic and non-White or Hispanic unmarried pregnant adolescents aged 12–17 years	1988–1989	Cross-sectional	Single time point	Individual
Gould et al. (1996)	–	Primary/secondary	267	New York, New Jersey, Connecticut	Black, White, and Hispanic children under age 20 years	1984–1986	Case-control	Retrospective	Individual
Hemphill et al. (2006)	–	Primary	3655	Washington, USA and Victoria, AUS <sup>a</sup>	Students in grade 7 and 9 aged 12–16 years <sup>b</sup>	2002 and 2003	Cohort	Prospective	Individual
Hemphill et al. (2012)	Social development model	Primary	3599	Washington, USA and Victoria, AUS <sup>a</sup>	White, Hispanic/Latino, Black, Native American, Asian/Pacific Islander, and other race students in grade 7 and grade 9	2002 and 2003	Cohort	Prospective	Individual
Kramer et al. (2017)	–	Primary	164	Boston, Massachusetts; New York, New York	White and non-White adolescents aged 13 through 17	Recruitment between 2007 and 2012	Case-control	Retrospective	Individual
Lee et al. (2014)	Jacobs' general theory of addiction; Social disorganization theory	Secondary	8318♦	Maryland	White, Black/African American, Hispanic/Latino, Asian/Pacific Islander, and other race students in grades 9 through 12 who engaged in at least one gambling activity in their lifetime	2012	Cross-sectional	Single time point	Multilevel
Paek et al. (2013)	Social ecological model	Secondary	983	Michigan	Black and White public school students in grades 9 through 12	2008 and 2009	Cohort	Prospective	Multilevel
Racz et al. (2016)	Deviance-proneness model	Primary	329♦	A secure, locked juvenile justice facility, California	Black, Latino, and White, boys experiencing incarceration, mean aged 16 years	Recruitment between 2005 and 2007	Cross-sectional	Single time point	Individual
Ramey (2016)	–	Secondary	3274	National	Black, Hispanic, and White boys, aged 0–14 at baseline (1988–2010)	1988–2010 and 1994–2012	Cohort	Prospective	Individual
Rosenbaum (2020)	Gottfredson and Hirschi self-control theory of deviance; Theory of labeling and deviance amplification	Secondary	1227♦	National	Propensity matched sample of Black, White, and Latino adolescents attending public and private middle, junior, and high schools	1995–2001	Cohort	Prospective	Individual
Rushton et al. (2002)	–	Secondary	~1248♦	National	Black, White, Hispanic, Asian, and other race adolescents in grades 7 through 12 at public and private middle, junior, and high schools mean aged 15.6 years with moderate/severe depression at baseline	1995–1996	Cohort	Prospective	Individual

(Continues)

**TABLE 2** (Continued)

Reference	Explicit guiding theory	Data source	Size (N)	Location	Population	Study period	Study design	Follow-up	Level of analysis
Slade (2004)	–	Secondary	2263	National	Black, Hispanic, and non-Hispanic White children aged 5 through 12 years old at baseline (1996)	1996 and 1998	Cohort	Prospective and single point in time	Individual
Talluri et al. (2014)	–	Secondary	1179	Houston, Texas	Nonsmoking, Mexican-American children aged 11 through 14	2005–2011	Cohort	Prospective	Individual

Note: Case–Control: participants selected to study based on outcome; Cohort: participants selected to study based on exposure; Cross-sectional: participants selected to study without regard to exposure or outcome; Prospective: exposure assessment precedes outcome assessment; Retrospective: outcome assessment precedes exposure assessment; Single point in time: exposure and outcome assessment at one time-point; Multilevel: group-level exposure and individual-level outcome with analysis accounting for clustering; Individual: exposure and outcome analyzed at the individual level.

Note: To reference descriptions of study participants’ demographic characteristics, we use language consistent with that of the authors. However, it is important to acknowledge that how these characteristics (e.g., race, ethnicity, and gender) are operationalized may vary across included articles.

<sup>a</sup>Included students from Victoria, Australia but adjusted for country (did not stratify analyses); N indicates total number of participants from both locales.

<sup>b</sup>Racial composition not reported.

♦ Authors report larger overall sample size; however, regression results presented are estimated in a subset of that sample.

Studies primarily examined more immediate onset health outcomes, with just four assessing outcomes in early- and/or mid-adulthood.<sup>61,62,67,69</sup> There was, however, a notable range in type of health outcomes examined, including mental (31.6%), behavioral (31.6%), and physical health (26.3%) or health systems involvement (10.5%) (Table 1). Mental health included measures of borderline personality disorder, depression, and persistent depressive symptoms, as well as outcomes whose categorization as mental health diagnoses has been critiqued.<sup>70</sup> These included adolescent antisocial behavior, adjustment problems, and pathological gambling. Behavioral health primarily focused on substance use (tobacco use, polysubstance use, and drug use disorder) or sexual activity (“risky sexual behavior”). Physical health included injuries, childbirth, asthma, sexually transmitted infection (STI), and death by suicide. Following guidance from the literature, we included death by suicide—as an indicator of mortality, with multicausal origins—in the physical health outcomes section.<sup>71,72</sup> Lastly, health system involvement included two measures of having received mental healthcare services. We elected not to collapse this category with the mental health category given that authors of these manuscripts explicitly differentiated between “receipt of” and “need for” services. They suggest receiving services may indicate either (1) a service need being met or (2) the unnecessary medicalization of a perceived misbehavior.<sup>61,73</sup> Overall, health indicators were primarily measured by student self-report (78.9%) using a binary questionnaire (yes/no) or by querying students on an index of symptoms (standardized scale administered by a researcher/clinician). Remaining studies used vital statistics/clinical data (medical test results and death certificate; 10.5%) or parental/caregiver report (10.5%) (Table 4).

## Findings

Among studies reporting precision estimates ( $n = 18$ ), most (61.1%) found experiencing punitive discipline was significantly associated with adverse health; the two studies (11.1%) examining health system involvement found punitive discipline was differentially associated with mental healthcare use by student racial identity; and the remaining studies (27.8%) reported null or nonsignificant results. Just over

half of the studies (52.6%) used either stratification or interaction to examine whether the school discipline–health association was modified by a third variable. Effect modifiers included race/ethnicity ( $n = 5$ ), gender ( $n = 3$ ), socioeconomic status, ( $n = 2$ ), geography ( $n = 2$ ), age ( $n = 1$ ), and grade ( $n = 1$ ). We present stratified results where available.

## Publications over time

Among the 19 studies comprising this literature, the earliest was published in 1990 (Figure 3). Collectively, the study periods in this literature provided coverage of the years 1966 through 2013. For 26.3% of the studies, follow-up fell within the “turn toward mandatory punishment” era (up to approximately the year 1990); over half (52.6%) were within the “height of mandatory punishment” era (approximately between the years 1990–2011); 10.5% were within the “considering alternatives to punishment” era (approximately the year 2012 to present); and 10.5% spanned multiple periods (Table 1). Older studies—that is, using data collected during the “turn toward mandatory punishment” era—primarily focused on physical health outcomes, while studies of data collected during the “height of mandatory punishment” era primarily focused on behavioral and mental health outcomes. Studies that reported null or nonsignificant results made up a larger proportion of the earlier “turn toward punishment” and later “considering alternatives to punishment” eras, and the “turn toward punishment” era studies also tended to have smaller samples (mean: 2655.8; median: 758; and range 241–10,362).

## Results of individual studies

### Mental health

Six studies examined associations between punitive school discipline and mental health, of which five found discipline increased risk and one found a null association. Studies that looked at differences by race suggested punitive school discipline was especially harmful to the mental health of Black students.

**TABLE 3** Results of individual included studies,  $N=19$ 

Reference	Discipline indicator	Health indicator	Effect measure modifiers	Covariates	Summary measures of association
Alexander et al. (1995)	School discipline problems	Injuries	Gender (stratified: male/female)	Main analyses: Gender Gender-stratified analyses: race, parent's education, exercise, employment, 30-day alcohol use, lifetime marijuana use, risk-taking, cruising, anger, parent supervision, BMI, pubertal status, and interactions between BMI and pubertal status	Main analyses RR: 2.08 ( $p < 0.05$ ) Stratified analyses: Male RR: 1.50 (95% CI: 1.19–2.07) Female RR: 1.42 (95% CI: 0.84–1.55)
Bottiani et al. (2017)	Black-White discipline gap	Adjustment problems	Race (stratified: Black/White)	Individual-level: gender, grade-level, maternal education level (proxy for SES); School-level: percent free and reduced-price meals, NGV school diversity statistics-, overall suspension rate, assigned randomization condition from original study	Stratified analyses: Black students standardized coefficient: 0.77 (SE: 0.17; $p < 0.001$ ) White students standardized coefficient: $-0.01$ (SE: 0.12; $p > 0.05$ )
Cottrell (2018)	Suspension	Depression (mediator)	Race (stratified: Black/White)	School attachment; college aspirations; deviant peer influence; peer substance use; social support; self-esteem; prior nonserious and serious delinquency; gender	Stratified analyses: High depression (vs. low): Black students OR: 1.89 (log odds: 0.64; SE: 0.28) White students OR: 1.41 (log odds: 0.34; SE: 0.19) Medium depression (vs. low): Black students OR: 0.92 (log odds: $-0.07$ ; SE: 0.26) White students OR: 0.84 (log odds: $-0.16$ ; SE: 0.18)
Elster et al. (1990)	School suspension	Age at first child's birth	Geography (stratified: urban/rural)	Race; age in 1980	Stratified: Urban: School-aged mothers versus young adult mothers OR: 1.56 School-aged mothers versus nonmothers OR: 2.44 Rural: School-aged mothers versus young adult mothers OR: 1.00 School-aged mothers versus nonmothers OR: 2.14 <i>Note: Authors do not present indicators of precision.</i>
Fothergill et al. (2008)	School suspension	Drug use disorder in adulthood	Gender (interaction term: male/female)	Gender; mother's self-report of drug use or regular alcohol use	Main analyses: Student self-report OR: 1.28 (95% CI: 1.07–1.52; $p < 0.01$ ) Maternal report OR: 1.04 (95% CI: 0.83–1.30; $p=0.73$ ) <i>Note: Tests for interaction found no significant interaction terms. Authors did not report estimates.</i>
Fowler et al. (1992)	History of suspension, exclusion, or expulsion	Current asthma; asthma severity	Family income (stratified: $<\$20k/\$20k+$ )	Current Asthma Model 1: child's sex, child's age group, maternal education level, family income, and race/ethnicity Current Asthma Model 2: child's sex, child's age group, maternal education level, family income, race/ethnicity, and categorical school days absent Stratified Models: maternal education, race/ethnicity, gender, age group Asthma Severity Models: child's gender, age group, maternal education, race/ethnicity	Main analyses (current asthma): Model 1 OR (comparing children with asthma to "group of well children"): 1.0 (95% CI: 0.6–1.6) Model 2 OR: 0.8 (95% CI: 0.5–1.4) <i>Note: Both models adjust for child's sex, age group; maternal education level; family income; and race/ethnicity. Model 2 additionally adjusts for days absent.</i> Stratified analyses: Family income $\geq \$20k$ OR: 0.6 (95% CI: 0.3–1.1) Family income $< \$20k$ OR: 1.2 (95% CI: 0.6–2.3) Main analyses (asthma severity; $N=536$ ): Health status measure OR: 1.3 (95% CI: 0.4–4.7) School absence measure (11–15 days) OR: 0.9 (95% CI: 0.2–4.6) School absence measure (16+ days) OR: 2.1 (95% CI: 0.6–7.4) Taking asthma medication measure OR: 1.2 (95% CI: 0.4–3.5) <i>Note: Despite the temporal ordering suggested by their measurement, all models regress history of suspension/expulsion on current asthma—effectively treating asthma as the exposure and discipline as the outcome.</i>
Gillmore (1992)	School problems	Risky sexual behavior	–	Model 1: years sexually active, age, race, family closeness, parental monitoring, lived with family, ran away from home, delinquent activities, partner used substances, general substance use Model 2: years sexually active, age, race, family closeness, parental monitoring, lived with family, ran away from home, delinquent activities, partner used substances, substance use during sex	Main analyses: Model 1 RD: $-0.02$ ( $p > 0.05$ ) Model 2 RD: $-0.05$ ( $p > 0.05$ )

(Continues)

**TABLE 3** (Continued)

Reference	Discipline indicator	Health indicator	Effect measure modifiers	Covariates	Summary measures of association
Gould et al. (1996)	Suspended from school	Suicide death	Gender (interaction term: male/female) Race (interaction term: African American/Hispanic/White/Other) Age (interaction term: aged 5–9, 10–14, 15–19) SES (interaction term: five-level categorical variable ranging from “the most impoverished” to “the most affluent class”)	Age; sex; ethnicity; Hollingshead’s Four-Factor Socioeconomic Status Index	Main analysis OR: 6.1 (95% CI: 1.6–23.4; $p < 0.01$ ) Note: Tests for interaction found no significant interaction terms, authors did not report estimates.
Hemphill et al. (2006)	School suspension (societal response to antisocial behavior)	Adolescent antisocial behavior	–	Demographic characteristics (female, age); individual risk factors (antisocial behavior, favorable attitude antisocial, favorable attitude drugs, attention/concentration problems, impulsivity); individual protective factors (religiosity, belief in the moral order, emotion control); family risk factors (poor family management, family conflict, parent favorable attitude to drugs, parent favorable attitude to antisocial); family protective factors (attachment to mother, attachment to father, opportunities for prosocial, recognition for prosocial); peer risk factors (association with antisocial peers); school risk factors (school grades); school protective factors (opportunities for prosocial, recognition for prosocial); community risk factors (low neighborhood attachment, community disorganization, perceived availability of drugs, norms favorable to drug use); community protective factor (opportunities for prosocial, recognition for prosocial); societal responses (arrests); state	Main analysis OR: 1.5 (95% CI: 1.1–2.1; $p < 0.05$ )
Hemphill et al. (2012)	School suspension (societal response to antisocial behavior)	Current tobacco use	Geography (interaction term: Victoria/Washington) Grade (stratified: grade 7/grade 9)	Demographic characteristics (gender, age, state); individual factors (current tobacco use, favorable attitudes to drug use, antisocial behavior, current alcohol use, current cannabis use); peer factors (interaction with antisocial peers, friends’ use of drugs); family factors (poor family management, conflict, parental attitudes favorable to drug use); school factors (academic failure, low commitment to school); community factors (norms favorable to drug use); risk factors (enforcement, availability of rugs); societal responses (arrests)	Stratified analyses: Grade 7 OR: 2.08 (95% CI: 1.23–3.52; $p < 0.01$ ) Grade 9 OR: 1.10 (95% CI: 0.66–1.81; $p > 0.05$ ) Note: Tests for interaction with geography found no significant interaction terms, authors did not report estimates.
Kramer et al. (2017)	Suspended or expelled	Borderline personality disorder (BPD)	–	Age	Main analysis OR: 8.12 (95% CI: 2.80–23.50; $p < 0.001$ )
Lee et al. (2014)	School-level student suspension rate	Pathological gambling	–	Student-level: age, male, African American, alcohol, cigarette, marijuana, nonmedical prescription drugs; School-level: student mobility rate, % student receiving free/reduced meals; % African American students; urban; rural; lifetime gambling prevalence; lifetime gambling problem prevalence; past-month substance use prevalence	Main analysis OR: 1.002 (95% CI: 0.998–1.005)
Paek et al. (2013)	Stringency of tobacco policy enforcement	Adolescent smoking behavior	–	Individual-level: sex, age, race; School-level: designation of a tobacco-free school zone, school smoking rate, other antismoking communications	Main analysis RD: –0.02 (SE: 0.01; $p=0.096$ )
Racz et al. (2016)	History of school expulsion	Polysubstance use	–	Race; age of onset of drug use; less than daily drug use; history of prior arrest	Main analysis IRR: 0.89 (95% CI: 0.79–1.01)

(Continues)

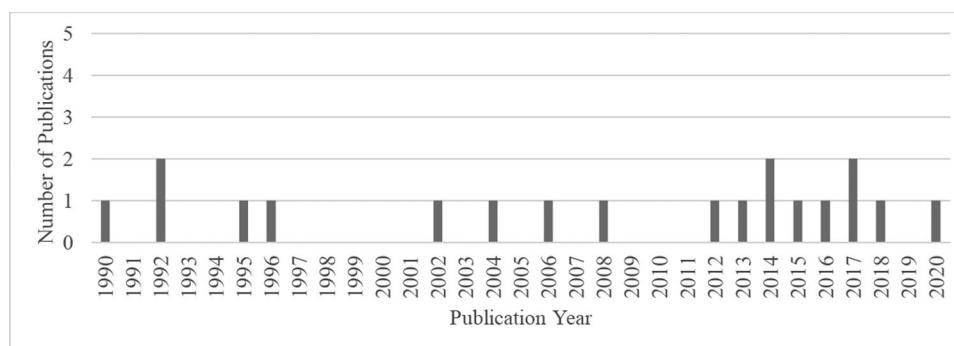
**TABLE 3** (Continued)

Reference	Discipline indicator	Health indicator	Effect measure modifiers	Covariates	Summary measures of association
Ramey (2016)	School punishment	Involvement in the mental health system	Race (stratified: Black/Hispanic/White)	Therapy/medication for behavior problems in childhood, Race (White, Black, Hispanic), childhood variables (externalizing behavior symptoms, PIAT math score, PIAT reading recognition score, repeated a grade, attended Head Start, poverty, year born), young adulthood variables (illegal activity, risk-averse attitudes, Rosenberg esteem scale score, depression, repeated a grade, graduated high school, income, economic idleness), residential status (suburban, rural, urban), region of country (Northeast, Midwest, South, West), insurance coverage (no insurance, public insurance, private insurance), age, age squared	Main analysis OR: 0.893 (log OR: $-0.114$ ; SE: 0.131; $p > 0.05$ ) Stratified analyses: Black students OR: 0.794 (log OR: $-0.231$ ; SE: 0.204; $p < 0.01$ ) Hispanic students OR: 0.699 (log OR: $-0.358$ ; SE: 0.279; $p > 0.05$ ) White students OR: 1.218 (log OR: 0.197; SE: 0.218; $p > 0.05$ )
Rosenbaum (2020)	Suspension	Chlamydia, trichomoniasis, and gonorrhea	–	Propensity matched on 67 covariates (demographics, socioeconomic status, sexual risk-taking, relationships with adults, educational factors, parents' risk behavior, substance use, personality and mental health, and deviance); for chlamydia and trichomoniasis analyses adjusted for baseline age, race/ethnicity, gender, and household income tertiles (no additional adjustment for gonorrhea analyses)	Matched analyses: Trichomoniasis OR: 2.87 (95% CI: 1.40–5.99; $p < 0.001$ ) Gonorrhea crude OR: 3.23 (95% CI: 0.95–11.03; $p=0.06$ ) Note: Suspended youth did not differ from nonsuspended youth in the adjusted odds of a positive chlamydia test in multivariate regression, authors did not report estimates.
Rushton et al. (2002)	Suspension	Persistent moderate/severe depressive symptoms	–	Race, grade in school, socioeconomic status, maternal educational status, single-parent household, received counseling, fair/poor general health, somatic symptoms, gender, unable to obtain needed medical care, suicidal ideation, family has fun together, close to father, family member completed suicide	Main analysis OR: 1.9 (95% CI: 1.3–2.7)
Slade (2004)	First school suspension or expulsion	Received mental health services	Race (stratified: Black/Hispanic/White)	Age, gender, BPI externalizing rating $\geq 125$ , BPI internalizing rating $> 124$ , child repeated grade during follow-up, child health fair or poor, mother's marital status, family income quartiles, health insurance coverage, mother's educational attainment, region, received counseling in year prior to baseline	Prebaseline first suspension/expulsion stratified analyses: Black Marginal RD: 0.011 ( $p=0.595$ ) Hispanic Marginal RD: 0.021 ( $p=0.593$ ) White Marginal RD: 0.077 ( $p=0.055$ ) Postbaseline first suspension/expulsion stratified analyses: Black Marginal RD: $-0.008$ ( $p=0.512$ ) Hispanic Marginal RD: $-0.010$ ( $p=0.803$ ) White Marginal RD: 0.209 ( $p < 0.001$ )
Talluri et al. (2014)	Detention or suspension	Smoking experimentation	–	Age, sex, cognitive susceptibility, tension, concentration, family cohesion, mother smoking, sister smoking, other smoking, peer influence, work smoking, neighborhood, thinking language, positive outcome expectations	Main analysis HR: 1.04 (beta: 0.036, SE: 0.015; $p=0.019$ )

Note: Where possible, we present 95% CI. If authors did not report 95% CI, we provide  $p$ -values. Elster et al. (1990) did not report indicator of precision.

Note: To reference descriptions of study participants' demographic characteristics, we use language consistent with that of the authors. However, it is important to acknowledge that how these characteristics (e.g., race, ethnicity, and gender) are operationalized may vary across included articles.

Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio; NGV, normalized generalized variance; OR, odds ratio; RD, risk difference; RR, risk ratio; SE, standard error; SES, socioeconomic status.

**FIGURE 3** Publications per year.

**TABLE 4** Measurement in individual included studies, N=19

Reference	Discipline indicator	Discipline operationalization	Health indicator	Health indicator operationalization
Alexander et al. (1995)	School discipline problems	Measurement: "Since the beginning of 8th grade, have you been in enough trouble to be sent to the principal's office?" (Student self-report: Yes/No)	Injuries	Measurement: "Report the number of injuries experienced in past year for which you received treatment by a physician or nurse?" (Student self-report: count data)
Bottiani et al. (2017)	Black-White discipline gap	Measurement: School discipline data available on the number of students with suspensions and total enrolled for the 2011–2012 school year disaggregated by race and gender (Secondary administrative data) Operationalization: Continuous variable: subtracted risk of suspension (risk = # of suspended students / # enrolled students) among White students from risk of suspension among Black students	Adjustment problems	Measurement: 4-item Adjustment Problems Scale—adapted from BASC-2 Externalizing Scale—measuring the frequency of externalizing symptoms: "I have trouble controlling my temper"; "I have threatened to hit or hurt someone"; "I do things without thinking"; "I get mad easily"; (Student self-report: 4-point Likert scale from "Never" to "Almost Always." Higher scores indicate higher levels of adjustment problems)
Cottrell (2018)	Suspension	Measurement: <i>Authors do not specify question;</i> (Student self-report: Yes/No)	Depression (mediator)	Measurement: 13-item scale; <i>Authors do not specify questions;</i> (Student self-report: <i>Authors do not specify response options</i> ) Operationalization: Three-level categorical variable: "low depression," "medium depression," and "high depression"
Elster et al. (1990)	School suspension	Measurement: "How many times during the past year [1980] did you experience school suspension?" (Student self-report: seven response categories ranging from "Never" to "more than 50 times") Operationalization: Dichotomous variable: Never suspended/Ever suspended	Age at first child's birth	Measurement: <i>Authors do not specify questions</i> (Student self-report: <i>Authors do not specify response options</i> ) Operationalization: Three-level categorical variable: "school-aged mothers" (i.e., 18 and younger at birth of the first child); "young adult mothers" (i.e., 19–21 years of age at birth of the first child); "nonmothers" (i.e., no child by age 21)
Fothergill et al. (2008)	School suspension	Measurement: "Please tell me how many times you've been suspended from school in the last 3 years?" (Student self-report and maternal report: "Never"/"Once"/"Twice"/"3 or 4 times"/"5 or more times") Note. Assessed in grade 7 and 8. Operationalization: Dichotomous: Never/Ever	Drug use disorder in adulthood	Measurement: Michigan version of Composite International Diagnostic Interview measured in early (age 32–33) and/or mid-adult (age 42–43) life; <i>Authors do not specify questions</i> (Student self-report: <i>Authors do not specify response options</i> ) Operationalization: Using computer algorithms of data to identify drug use disorders
Fowler et al. (1992)	History of suspension, exclusion, or expulsion	Measurement: Has the study child "ever been suspended, excluded, or expelled from school?" (Parental/caregiver-report: Yes/No)	Current asthma; asthma severity	Current asthma measurement: "Has the child ever had asthma?" and "has the child had asthma in the past 12 months?" (Parental/caregiver-report: Yes/No) Current asthma operationalization: Measures were combined such that only children who had asthma in the past year were considered as currently having asthma Asthma severity measurement: From 1988 NHIS-CH survey: "Health status" (Parental/caregiver-report: Fair or poor/Excellent to good); "Use of asthma medication" (Parental/caregiver-report: yes/no); "During the past 12 months, that is since a year ago, about how many days was child absent from school because of illness?" (Parental/caregiver-report: 0–10 days/11–15 days/16+ days)
Gillmore (1992)	School problems	Measurement: Number of times the respondent had ever been suspended or expelled (Student self-report: <i>Authors do not specify response options</i> )	Risky sexual behavior	Measurement: Number of the following in which the respondent has ever engaged: anal sex, sex in exchange for money, sex in exchange for drugs, more than one sexual partner, and sex with a casual partner; (Student self-report: "never engaged"/"ever engaged"). Operationalization: Responses totaled, forming scale ranging from 0 = no sexual risk-taking to 5 = engaging in five different types of risky sexual behavior
Gould et al. (1996)	Suspended from school	Measurement: Having been suspended within 3 months of death or assessment (Student self-report or report of sibling/friend and caregiver and one to three school teachers: Yes/No)	Completed suicide	Measurement: Documented suicides completed by persons younger than 20 in 2-year period ( <i>Authors do not specify source of data, likely death certificate data</i> )

(Continues)

**TABLE 4** (Continued)

Reference	Discipline indicator	Discipline operationalization	Health indicator	Health indicator operationalization
Hemphill et al. (2006)	School suspensions (societal response to antisocial behavior)	Measurement: "How many times in the past year have you been suspended from school?" (Student self-report: 8-point Likert scale, ranging from "Never" to "40 or more times") Operationalization: Dichotomous variable: Suspension 1+ times in the past year/did not experience suspension	Adolescent antisocial behavior	Measurement: "How often in past year have you: carried a weapon; stolen something worth more than \$5 (U.S.) or \$10 (Australia); attacked someone with the idea of seriously hurting them; sold illegal drugs; stole or tried to steal a motor vehicle such as a car or motorcycle; been drunk or high at school; taken a handgun to school; threatened someone with a weapon; beaten up someone so badly that s/he required medical treatment" (Student self-report: ranged from "Never" to "40 or more times" on an 8-point scale) and "How many whole days during the last four weeks have you missed school because you skipped or wagged (i.e., engaged in truancy)?" (Student self-report: rated from "None" to "11 or more") Operationalization: Dichotomized variable into present (students engaged in the behaviors one or more times in the past year)/absent (students never engaged in the behaviors listed). Then, created index of antisocial behavior by summing counts of each present behavior. Possible range: 0–9 (where 0–2 = "Nonantisocial Group" and 3–9 = "Antisocial Group")
Hemphill et al. (2012)	School suspensions (societal response to antisocial behavior)	Measurement: "How many times in the past year have you been suspended from school?" (Student self-report: 8-point Likert scale, ranging from "Never" to "40 or more times") Operationalization: Dichotomous variable: Suspension 1+ times in the past year/did not experience suspension	Current tobacco use	Measurement: "How frequently have you smoked cigarettes in the past 30 days?" (Student self-report: 8-point Likert scale ranging from "Not at all" to "40 or more per day"). Operationalization: Dichotomous variable: Never or no use/Other than never or none
Kramer et al. (2017)	Suspended or expelled	Measurement: Adolescent version of Background Information Schedule assessing psychosocial functioning in the past 2 years. Assessed under "vocational functioning." <i>Authors do not specify how question was posed</i> (Student self-report: <i>Authors do not specify response options</i> )	Borderline personality disorder (BPD)	Measurement: Meeting both Revised Diagnostic Interview for Borderlines and DSM-IV criteria for BPD; <i>Authors do not specify questions</i> (Student self-report: <i>Authors do not specify responses</i> ). Operationalization: Dichotomous variable: psychiatrically healthy adolescents/adolescent BPD
Lee et al. (2014)	School-level student suspension rate	Measurement: Total number of suspensions; total student enrollment (Secondary administrative data from Maryland State Department of Education) Operationalization: Divided the total number of suspension incidents by total student enrollment	Pathological gambling	Measurement: Two-item (selected from DSM-IV criteria for pathological gambling) on Lie/Bet Questionnaire "Have you ever felt the need to bet more and more?" and "Have you ever had to lie to people important to you about how much you gambled?" (Student self-report: yes/no). Operationalization: Dichotomous variable: no gambling problems/any gambling problems
Paek et al. (2013)	Stringency of tobacco policy enforcement	Measurement: "When students are caught smoking cigarettes, how often is each of the following actions taken? (1) placed in detention; (2) not allowed in extra-curricular activities; (3) given in-school suspension; (4) suspended from school; (5) expelled from school; and (6) reassigned to alternative school." (Administrator reports: 4-point ordinal scale including Sometimes/Always/Almost Always/Never or rarely) Operationalization: Dichotomous variable: Yes (Sometimes, always, almost always)/No (Never or rarely). Then, "yes" response was counted.	Adolescent smoking behavior	Measurement: "During the past 30 days, on how many days did you smoke cigarettes?" (Student self-report: 0 days/1 or 2 days/3–5 days/6–9 days/10–19 days/20–29 days/all 30 days); "During the past 30 days, on the days you smoked, how many cigarettes did you smoke per day?" (Student self-report: None/Less than 1 cigarette/1 cigarette/2–4 cigarettes/6–10 cigarettes/11–20 cigarettes/More than 20 cigarettes) Operationalization: Averaging index of both questions due to high correlation between these measures ( $r=0.931$ , $p < 0.001$ ).

(Continues)



**TABLE 4** (Continued)

Reference	Discipline indicator	Discipline operationalization	Health indicator	Health indicator operationalization
Racz et al. (2016)	History of school expulsion	Measurement: "Have you experienced an expulsion from school in the past?" (Student self-report: Yes/No)	Polysubstance use	Measurement: 22-item drug use scale where interview questions included: "How often have you used this substance?" "How old were you when you first tried this substance?" "How often have you used in the past 6 months?" regarding the following substances: alcohol, marijuana, inhalants, and "any other drugs" (Youth self-report: open-response and count) Operationalization: Count of substances each participant endorsed using over lifetime. Higher scores indicative of polysubstance use.
Ramey (2016)	School punishment	Measurement: "Has your child ever been suspended or expelled from school?" (Maternal report: Yes/No)	Involvement in the mental health system	Measurement: "Have you seen a psychiatrist or psychologist for troubles in school, a violent temper, or behavior problems at school or work since the last interview?" "Have you taken drugs to control behavior since the last interview?" in early adulthood (age 15–35) (Student self-report: Yes/No) Operationalization: Dichotomous variable: Yes (if respondent answered "yes" to one or both questions)/No
Rosenbaum (2020)	Suspension	Measurement: "During this school year did you receive an out-of-school suspension from school?" (Student self-report: Yes/No). <i>Note: Since sample was limited to participants without prior out-of-school suspension or expulsion, affirmative response indicated first lifetime suspension.</i>	Chlamydia; trichomoniasis; and gonorrhea	Measurement: Chlamydia and gonorrhea detected with Ligase ChainReaction amplification technology in the Abbot LCxProbe System. Trichomoniasis detected with PCR-ELISA test in early adulthood (age 18–25) (Clinical data; Testing positive for Chlamydia trachomatis, Neisseria gonorrhoeae, or Trichomonas vaginalis)
Rushton et al. (2002)	Suspension	Measurement: <i>Authors do not specify how question was posed</i> (Student self-report: <i>Authors do not specify response options</i> ) Operationalization: Dichotomous variable: Yes/No	Persistent moderate/severe depressive symptoms	Measurement: Frequency of depressive symptoms in the past week based on 20 AddHealth survey items corresponding to 20-item CES-D with slight modifications to items on symptoms of "restless sleep" and "crying spells"; (Student self-report: ranged from 0 = "Never" to 3 = "Daily") Operationalization: Total CES-D scores represent summation of 20 items, producing possible range from 0 = "no depressive symptoms" to 60 = "most frequent/severe depressive symptoms." Categorized depressive symptoms using adult cutoff of 16 and Roberts' adolescent cutoff of 24 such that 0–15 = "minimal depressive symptoms"; 16–23 = "mild depressive symptoms"; ≥24 = "moderate/severe depressive symptoms."
Slade (2004)	First school suspension or expulsion	Measurement: "Has your child ever been suspended or expelled from school?" (Maternal report: Yes/No) Operationalization: Postbaseline first suspension or expulsion: "No" at baseline and "Yes" at time point 2 indicated that student had been suspended for first time during the period between interviews Prebaseline first suspension or expulsion: "Yes" at baseline indicated that student had been suspended or expelled for the first time prebaseline interview <i>Note: It was not possible to determine if students suspended/expelled prebaseline were suspended/expelled again postbaseline.</i>	Received mental health services	Measurement: "Did your child see a psychiatrist, psychologist, or counselor for a mental or emotional problem in the past 12 months?" (Maternal report: Yes/No)
Talluri et al. (2014)	Detention or suspension	Measurement: "During this school year how many detentions or suspensions have you had?" (Student self-report: <i>Authors do not specify response options</i> ) Operationalization: Dichotomous variable: None/One or more	Smoking experimentation	Measurement: "Have you ever smoked a cigarette?" and "Have you ever tried a cigarette, even a puff?" (Student self-report: Yes/No) Operationalization: Dichotomous Variable: Nonexperimenters ("No" to both questions)/Experimenters ("Yes" to either question)

Abbreviations: BASC-2, Behavioral Assessment System for Children: Second Edition; CES-D, Center for Epidemiological Studies-Depression; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, 4th Edition; NHIS-CH, National Health Interview Survey on Child Health; PCR-ELISA, polymerase chain reaction–enzyme-linked immunosorbent assay.

Specifically, experiencing punitive school discipline was associated with a higher odds of persistent depressive symptoms (OR: 1.9; 95% CI: 1.3–2.7;  $N = 1248$ ; individual-level analysis);<sup>74</sup> high depression among Black students (OR: 1.89;  $p < 0.05$ ) but not White students (OR: 1.41;  $p > 0.05$ ) ( $N = 3289$ ; individual-level analysis);<sup>75</sup> and borderline personality disorder (OR: 8.12; 95% CI: 2.80–23.50;  $N = 164$ ; individual-level analysis).<sup>76</sup> Similar patterns presented among studies examining outcomes whose categorization as mental health diagnoses has been critiqued (i.e., adjustment problems, antisocial behavior, and pathological gambling). Namely, school-level Black-White suspension gap was associated with higher levels of adjustment problems for Black students (standardized coefficient: 0.77;  $p < 0.001$ ) but not White students (standardized coefficient:  $-0.01$ ;  $p > 0.05$ ) ( $N = 19,726$ ; multi-level analysis)<sup>64</sup> and experiencing suspension was associated with higher odds of antisocial behavior (OR: 1.5; 95% CI: 1.1–2.1;  $N = 3655$ ; individual-level analysis).<sup>77</sup> Just one study found a null association: school-level student suspension rate was not significantly associated with pathological gambling (OR: 1.002; 95% CI: 0.998–1.005;  $N = 8318$ ; multi-level analysis).<sup>65</sup>

## Behavioral health

The six studies assessing behavioral health outcomes yielded mixed results. Three examined smoking-related indicators, finding that punitive school discipline was associated with an increase in the 5-year hazard of smoking experimentation (HR: 1.04; 95% CI: 1.01–1.07;  $p = 0.019$ ;  $N = 1179$ ; individual-level analysis)<sup>78</sup> and a higher odds of current tobacco use among seventh grade students (OR: 2.08; 95% CI: 1.23–3.52) but not ninth grade students (OR: 1.10; 95% CI: 0.66–1.81;  $N = 3599$ ; individual-level analysis).<sup>79</sup> One study found no significant association between tobacco policy enforcement stringency and 30-day smoking behaviors at the 0.05-level (RD =  $-0.02$ ;  $p = 0.096$ ;  $N = 983$ ; multi-level analysis).<sup>66</sup>

Other substance use outcomes focused on drug use disorder and polysubstance use (number of different substance types used). One study found suspension was associated with higher odds of drug use disorder in adulthood (OR: 1.28; 95% CI: 1.07–1.52) when the exposure was measured using student reports, but when measured with maternal reports, the association was attenuated and no longer significant (OR: 1.04; 95% CI: 0.83–1.30;  $N = 1053$ ; individual-level analysis).<sup>67</sup> Another study among incarcerated boys found no significant association between punitive school discipline and polysubstance use (IRR: 0.89; 95% CI: 0.79–1.01;  $N = 329$ ; individual-level analysis).<sup>60</sup>

Lastly, one study<sup>63</sup> of pregnant adolescents examined whether punitive school discipline was associated with “risky sexual behaviors” (see Table 4 for operationalization). They found no significant association between the number of times participants had ever been suspended/expelled and the number of “risky sexual behaviors” in which they participated ( $N = 241$ ; individual-level analysis).

## Physical health

Five studies examined associations between punitive school discipline and physical health, four reporting an increased risk and two reporting null or nonsignificant associations. Specifically, punitive school discipline was associated with higher odds of suicide death (OR: 6.1; 95% CI: 1.6–23.4;  $N = 267$ ; individual-level analysis)<sup>68</sup> and higher odds of trichomoniasis at 5-year follow-up (OR: 2.9; 95% CI: 1.40–5.99)<sup>69</sup> but had no significant association with gonorrhea (unadjusted OR: 3.23; 95% CI: 0.95–11.03) or chlamydia (estimates not reported) ( $N = 1227$ ; individual-level analysis). Rosenbaum also tested 34 postsuspension variables as potential mediators finding none significantly ( $p < 0.05$ ) mediated the suspension-trichomoniasis association. Another study found being sent to the principal's office was associated with later injury for which participants sought clinical treatment (RR: 2.08;  $p < 0.05$ ;  $N = 758$ ; individual-level analysis), positing that experiencing punitive discipline may indicate a propensity toward risk-taking behaviors that result in injury.<sup>80</sup> When they stratified their analysis by binary gender, results were significant for boys (RR: 1.50; 95% CI: 1.19–2.07) and not girls (RR: 1.42; 95% CI: 0.84–1.55). Another study among teenage girls found experiencing suspension was associated with higher odds of having a first child at age 18 or younger ( $N = 1651$ ; individual-level analysis). The authors did not present indicators of precision around their estimates.<sup>62</sup> Lastly, one study found no significant association between punitive school discipline and either current asthma (OR: 0.8; 95% CI: 0.5–1.4) or asthma severity (OR: 1.2; 95% CI: 0.4–3.5;  $N = 10,362$ ; individual-level analysis)<sup>81</sup> with results persisting when stratified by family income.

## Health system involvement

Two studies explicitly assessed indicators of health systems involvement, both of which focused on mental health services. One found punitive discipline was associated with greater use of mental health services among White students (Marginal RD: 0.209;  $p < 0.001$ ) with no significant association among Black (Marginal RD:  $-0.008$ ;  $p = 0.512$ ) or Hispanic (Marginal RD:  $-0.010$ ;  $p = 0.803$ ) students ( $N = 2263$ ; individual-level analysis).<sup>82</sup> While the other study found no statistically significant overall punitive discipline-mental health system involvement association (OR: 0.893;  $p > 0.05$ ), it similarly documented racial disparity such that punitive school discipline was statistically significantly associated with lower odds of mental health system involvement in early adulthood for Black students (OR: 0.794;  $p < 0.01$ ) with no significant relationship for Hispanic (OR: 0.699;  $p > 0.05$ ) or White (OR: 1.218;  $p > 0.05$ ) students ( $N = 3274$ ; individual-level analysis).<sup>61</sup>

## DISCUSSION

This systematic review yielded 19 studies on the punitive school discipline–health association. It suggests that punitive school discipline

in early life is largely a harmful exposure adversely impacting health. Specifically, punitive school discipline was associated with an increased risk for: persistent depressive symptoms, depression, drug use disorder in adulthood, borderline personality disorder, antisocial behavior, death by suicide, injuries, trichomoniasis, pregnancy in adolescence, tobacco use, and smoking experimentation. Studies also documented differential impacts by race, gender, and grade. Studies reported null or nonsignificant results for outcomes of: pathological gambling, polysubstance use, “risky sexual behaviors,” gonorrhea, chlamydia, and asthma. Additionally, one study of tobacco policy enforcement stringency found no statistically significant association with 30-day smoking behavior. Particularly given the prevalence and inequitable distribution of exposure to punitive discipline in US schools, these findings indicate that punitive school discipline may play a role in patterns of population health and health inequity over the lifecourse.

## Revisiting the conceptual framework

### Material, psychosocial, and behavioral pathways

While the adapted framework suggests several complex and interconnecting factors that link exposure to punitive school discipline with poorer health, only one study empirically tested underlying mechanisms. Still, overall, findings from this review are consistent with what we would expect to observe given the framework. For example, loss of material resources via school exclusion—ranging from fundamental needs for wellbeing, like daily meals, to those promoting educational advancement, like access to curriculum<sup>83–85</sup>—may shape short- and long-term physical and mental health through myriad mechanisms, including food insecurity<sup>83,86</sup> and/or educational trajectory disruption. Only one study tested for mediation, finding no significant evidence at the 0.05-level that the association between punitive discipline and trichomoniasis in early adulthood was mediated by having received a high school diploma.<sup>69</sup> However, given educational attainment is largely associated with health through cumulative mechanisms (e.g., future employment and income), it may be attainment more likely mediates chronic health outcomes in later life.<sup>1,46,48,49</sup> Alternatively, examining whether disruption itself is health-harming, independent of its effects on attainment, may be worthwhile.

Studies tying punitive discipline to behavioral health (e.g., current tobacco use) also aligned with pathways proposed in our adapted conceptual framework.<sup>78,79</sup> Various hypotheses (e.g., coping behaviors) explain similar observations in a broader literature connecting negative early-life events—like isolation and emotional neglect—to increased drug use, misuse, and addiction.<sup>50,87</sup> Indeed, the loss of supportive school relationships following punitive discipline (e.g., school disconnectedness) may exacerbate these processes.<sup>52,88</sup> One study, however, found no significant change in adolescent smoking behaviors as tobacco policy enforcement stringency increased.<sup>66</sup> Together, these findings may indicate that if punishment increases the risk of smoking experimentation and current tobacco use, then when

leveraged to deter smoking behaviors, it may instead result in their persistence.<sup>66,78,79</sup>

Behavioral and psychosocial pathways may also explain physical health outcomes, such as death by suicide<sup>68</sup> or injury.<sup>80</sup> For example, adverse life experiences in youth, like punitive school discipline, comprise stress-related adversities that increase the risk for both (1) compounding experiences, like legal system targeting<sup>54</sup> which can directly cause injury,<sup>89</sup> and (2) outcomes like depression and depressive symptoms.<sup>74,75,90,91</sup> Depression and other stress-related mental health outcomes are associated with suicide death<sup>92</sup> and injury (e.g., due to self-, interpersonal-, and/or structural-harm).<sup>93–95</sup> Future research examining these proposed pathways is warranted.

### Punitive discipline and lifecourse health

Consistent with our adapted framework and lifecourse theory, studies in this review suggest that there may be sensitive time periods during which exposure to punitive discipline is particularly health-harming.<sup>2,38</sup> Most studies focused on students enrolled in high school and those that included students in other grades tended to estimate average associations over the multigrade sample. One study, however, documented differential findings when stratified by grade: the odds of tobacco use doubled for seventh-grade students who experienced suspension, with no significant effects among ninth-grade students.<sup>79</sup> This may indicate that middle school is a time during which school discipline increases the risk of initiating new tobacco use practices, whereas, by high school, practices may be more established and thus less affected by school discipline.<sup>96</sup> Studies examining the consistency of this observation are needed.

Adverse implications of exposure during sensitive periods may also persist into adulthood. Lifecourse theory and prior research have attributed this, in part, to changes in development, leading to physiologic dysregulation, accelerated aging, and reductions in coping, all expected responses to an external stressor.<sup>2,38,50,87,90</sup> Given punitive discipline excludes students from their classrooms/schools, lifecourse mechanisms may also include cumulative material resource loss via sequelae of earlier onset and/or repeated educational disruption—what lifecourse theory calls “the accumulation of risk.”<sup>97</sup> Four studies in this review examined exposure in early-life and outcomes in early/mid-adulthood, linking punitive school discipline to higher odds of having a child at younger ages;<sup>62</sup> trichomoniasis diagnosis in early adulthood;<sup>69</sup> lower odds of involvement with the mental health system in adulthood among Black men;<sup>61</sup> and higher odds of drug use disorder in early (age 32–33) and/or mid-adulthood (age 42–43),<sup>67</sup> the magnitude of which was larger than that of substance use outcomes estimated cross-sectionally, potentially further evidence of latent effects.

Together, these lifecourse findings suggest that middle school could be an especially sensitive time period, whereby the health effects of exposure to punitive discipline are compounded immediately and into adulthood. Future work should examine the persistence of this pattern.

## Societal ideologies, values, governance, and schools

With *Societal Ideologies* enveloping the framework, this indicates that the proposed mechanisms (from normalizing *Societal Values* around punishment as a means of accountability, justice, or control; to punitive policies at federal, state, and local levels of *Governance*; to *School* environments that attribute perceived student misbehavior to inherent characteristics of the student, their peers, or communities; to material, psychosocial, and behavioral pathways) are imbedded with inequity, and if left unchecked, produce disproportionalities in both exposure to punitive discipline and associated outcomes.<sup>5,31,33–36,42</sup> Findings in this literature are consistent with this overall structure.

Indeed, this review included studies documenting a disproportionately greater risk of health outcomes like depression, among Black students compared with White students.<sup>75</sup> These findings are consistent with the public health literature on discrimination and health, which documents how discriminatory policies and practices that deny structurally marginalized people access to full educational, economic, occupational, residential, and political resources, in turn, constrain opportunities for moving about in healthful environments.<sup>98</sup> Further, that literature describes discrimination as a stressor adversely affecting psychological wellbeing and health behaviors, with implications for physical and mental health.<sup>98</sup> By extension, this broader literature provides support for the physiologic plausibility of the association between discriminatory school discipline and health (i.e., health sequelae of denying access to educational resources and/or exposure to a racialized stressor).

The structure of this adapted framework also helps to interpret results from the two studies on mental health services use, with one finding punitive school discipline increased mental health service use among White boys and another finding it decreased mental health service use among Black boys.<sup>61,82</sup> Reflecting similar patterns in other institutions (e.g., the legal system), this literature attributes these inequities to societal ascriptions of “blameworthiness,” and “perceptions of threat,” for Black children versus underlying “biological or psychological causes” for White children.<sup>99,100</sup> In sum, whereas White children are more likely to have a medicalization frame deployed to respond to perceived misbehavior in schools, Black children are more likely to have a criminalization frame deployed.<sup>39</sup> While both approaches constitute means of social control for perceived inherent attributes, their nature and lifecourse consequences differ.

## Limitations

### Limitations of this literature

There are limitations to this literature that ought to be considered when interpreting results. First, over one-third of the studies assessed exposure and outcome at a single time point, posing potential challenges for establishing the directionality of estimated associations.

Second, given an emphasis on inherent/interpersonal theoretical frameworks, the sample researchers selected for some studies may

have been analytically limited in documenting associations between punitive school discipline and health. For example, in the study examining whether the history of school expulsion was associated with polysubstance use among young people experiencing incarceration,<sup>60</sup> given all participants were simultaneously exposed to a punitive system, the underlying mechanisms driving associations between punitive school discipline and health likely did not vary between exposed and unexposed groups, possibly explaining the null results.

Third, there was notable variability in the measurement of punitive discipline. Studies used both individual- (e.g., first suspension; lifetime suspension; number of times suspended) and school-level (e.g., Black-White suspension gap) measures with assessment at different developmental time periods and from different informants. For example, one study measured suspension using both student and maternal reports, with parents less likely to report that their child received suspensions from school.<sup>67</sup> Whereas student suspension report was associated with higher odds of drug use disorder in adulthood, associations between maternal report of child suspension and later life drug use disorder were not statistically significant. While evidence supports that student reports may be a better measure of school-based experiences—and particularly of how students internalize those experiences—there may also be distinct and important elements that maternal report is capturing (e.g., parental notification of child school discipline experiences).<sup>101</sup> Ultimately, each of the above approaches could be measuring different elements of a broader punitive discipline construct (e.g., Black-White suspension gap more explicitly capturing a racial discrimination domain), with implications for which health outcomes we may observe changes in, when we assess for health effects, and among whom.

Fourth, while each study adjusted for potential confounders, the robustness of adjustment varied by study (Table 3). Particularly among studies that minimally adjusted for potential confounders—residual or unmeasured confounding may result in biased estimates. Several studies, however, endeavored to eliminate alternative explanations by adjusting for a battery of covariates. In these studies, the association between punitive discipline and increased risk for their respective health outcomes remained after adjustment.

Fifth, half of the studies that reported null or nonsignificant findings for their overall analyses had samples comprised of fewer than 1000 participants. Further, studies that performed stratified analyses reported significant results for some groups of participants and not others. If adequately powered, these findings could indicate that punitive school discipline has no overall effect on the particular health outcomes under study in these analyses and, in the latter case, serve as evidence of effect modification. A possible alternative explanation could be that some of these studies were underpowered to detect statistical differences, overall or within certain populations. For example, in their overall analytic sample of 1227 participants, Rosenbaum found punitive school discipline was associated with a statistically significantly higher risk for the STI outcome of trichomoniasis (OR: 2.9; 95% CI: 1.40–5.99) but reported nonsignificant results for gonorrhea (unadjusted OR: 3.23; 95% CI: 0.95–11.03) and chlamydia (estimates not provided). As the author notes, however, their sample

varied by outcome, with the greatest missingness for gonorrhea ( $n = 810$  participants missing data) which was also the rarest outcome ( $n = 21$  cases)—all of which could have implications for precision around the estimate. Confirmatory analysis among estimates with notably wide confidence intervals is warranted.

Lastly, one prospective cohort study appeared to adjust for post-exposure covariates which could bias estimates as these factors may mediate the exposure–outcome relationship.<sup>102</sup> If so, the reported association (i.e., first-time school suspension and receipt of mental health services) may be an underestimation of the true association.

## Limitations of our review

There were also limitations to our review approach. First, there were some studies that appeared to conceptualize the temporality of their exposure and outcome differently from how they operationalized those relationships. For example, one study motivated their approach and interpreted their results with punitive school discipline as the exposure and health (age at first child's birth) as the outcome.<sup>62</sup> Their measurement of punitive discipline accordingly preceded that of age at the first child's birth. In their analysis and results, however, they position age at the first child's birth as the exposure and history of punitive school discipline as the outcome. Across the four studies for which a similar situation occurred,<sup>62,65,76,81</sup> punitive discipline appeared to precede the health indicators so we included them. However, inclusion in this literature on punitive discipline as a cause and health as an outcome should be interpreted with caution. Second, two studies included a sample of Australian students in their study population.<sup>77,79</sup> We retained these analyses as they adjusted for the state (Washington, US; Victoria, Australia), but again urge caution in interpretation within this US-specific literature.

Third, despite our rigorous selection strategy, it is possible we missed relevant studies. For example, while our health search terms included “health, morbidity, mortality, health disparities, health inequities, health status, wellbeing,” we did not search for particular conditions (e.g., diabetes). Therefore, if studies exclusively use condition-specific language in their title and abstract, they may have been excluded from this analysis. This would not, however, be the case for our PubMed search as we excluded health-specific terms from our string when searching this database.

Finally, as with all reviews, there is the potential for publication and reporting bias. That is, given null and nonsignificant results are less likely to be presented or published, there could be an overrepresentation of studies with statistically significant findings and/or those that report associations that are larger in magnitude.

## Implications for future research

First, with 58 articles excluded at Stage 3 of our search for examining discipline as an outcome, it appears the broader literature primarily conceptualizes punitive discipline as a health consequence rather than

a cause. We note this relationship is likely cyclical, whereby students who are punitively disciplined may experience, for example, adverse mental health outcomes, and students with mental health conditions are at increased risk for punitive discipline. Future studies should examine discipline as an exposure to better understand its implications for health and/or examine the cyclicity of this relationship. Further, given several studies measured the exposure and outcome at a single time point with potential implications for establishing temporality, future research should examine these associations in longitudinal datasets with prospective follow-up.

Second, just one study empirically examined potential mechanisms linking punitive school discipline to health and few studies examined outcomes in adulthood. Analyses should assess potential mediators of the association between exposure to punitive discipline in early life and health outcomes over the lifecourse. Our adapted framework may be useful for informing the development of testable hypotheses.

Third, to account for alternative explanations for the association between punitive school discipline and later health, future work should continue to build from the compelling confounder adjustment approaches implemented by more recent studies within this literature (i.e., matching students with variable school discipline experiences; matching students across schools with variable school discipline policies) as well as other design approaches in the broader punitive school discipline literature (i.e., natural experiments using changes in school district boundaries between districts with different propensities to discipline students).<sup>103</sup>

Further, most analyses were performed at the individual-level, foregoing analysis of punitive discipline as a systems-level determinant of health and limiting the possibilities for structural intervention. Future research should position school discipline policy at national, state, and local levels as primary exposures to document their association with young people's health.<sup>9</sup> Moreover, leveraging large, longitudinal datasets to examine the health effects of changes in school discipline policy over time may also facilitate a better understanding of whether the null or imprecisely estimated results observed for outcomes primarily examined within particular time periods are (1) consistently null over time (indicating school discipline policy does not impact these outcomes); (2) attributable to issues around statistical power (indicating prior studies were underpowered to detect differences); and/or (3) a function of changing policy (indicating particular types of policies that increase or decrease risk, and the latency of effects).

Finally, while some studies examined variation in the association across experiences of marginalization, few invoked theory to support these analyses and none considered intersectional axes of oppression, which may obscure important patterns. For example, while the prevalence of punitive discipline is highest among boys, emerging research has demonstrated that, consistent with population patterns of racialized targeting for policing and incarceration of Black women, at the intersections of racial and gender identity, Black girls are disproportionately targeted for punitive discipline.<sup>9,12,104</sup> Work examining the effects of this on the health experiences of Black girls is warranted. Further, absent from this literature is an explicit examination of how other experiences of structural marginalization—which are also



known to be disproportionately targeted for punitive discipline—may be uniquely impacted. For example, additional research examining implications for Native/Indigenous, LGBTQIA, and disabled young people is critical. Essential to pursuing this inquiry is the explicit use of theory (e.g., critical race theory) positioned to examine the complex ways structural marginalization (e.g., racism and ableism) may operate through punitive school discipline to shape inequitable distributions of health outcomes.

## CONCLUSION

Experiences of punitive discipline are common among young people attending US schools with Black, Latinx, Native/Indigenous, LGBTQIA, and disabled students disproportionately targeted. We conducted a systematic review on the health implications of punitive discipline in US schools, finding that evidence to-date suggests punitive school discipline is largely a health-harming exposure, with implications for racial health inequity. Together with research on the adverse effects of punitive school discipline on educational and legal system outcomes, this analysis adds support to demands from students, organizers, and others to move away from punitive discipline toward health-affirming interventions to ensure school connectedness, safety, and wellbeing. As legislators at federal, state, and local levels revisit related policies, consideration of these public health implications is critical.

## AUTHOR CONTRIBUTIONS

C.d.P.D. conceptualized the manuscript and the adapted conceptual framework. C.d.P.D., C.M., M.B., S.K., and S.J.P. collected and analyzed the data. C.d.P.D. interpreted the data and drafted the manuscript. J.S. and M.S.M. provided supervision and feedback throughout the process. All authors provided critical review and approved the final submission.

## ACKNOWLEDGMENTS

We would like to thank Jennifer Ahern and Anusha Vable for their critical feedback on this review, and Michael Sholinbeck for his support and expertise in generating our search strategy. C.d.P.D. was supported by a Health Policy Research Scholars program grant from the Robert Wood Johnson Foundation.

## COMPETING INTERESTS

The authors have no competing interests to declare.

## REFERENCES

- Walsemann, K. M., Gee, G. C., & Ro, A. (2013). Educational attainment in the context of social inequality: New directions for research on education and health. *American Behavioral Scientist*, 57(8), 1082–1104.
- Glymour, M. M., Avendano, M., & Kawachi, I. (2014). Socioeconomic status and health. In L. F. Berkman, I. Kawachi, & M. M. Glymour (Eds.), *Social epidemiology* (pp. 17–63). Oxford University Press.
- Bonell, C., Parry, W., Wells, H., Jamal, F., Fletcher, A., Harden, A., Thomas, J., Campbell, R., Petticrew, M., Murphy, S., Whitehead, M., & Moore, L. (2013). The effects of the school environment on student health: A systematic review of multi-level studies. *Health & Place*, 21, 180–191.
- Ancheta, A. J., Bruzzese, J. M., & Hughes, T. L. (2021). The impact of positive school climate on suicidality and mental health among LGBTQ adolescents: A systematic review. *Journal of School Nursing*, 37(2), 75–86.
- Punitive vs. restorative approach to school discipline. (2021). [https://www.restorativeresources.org/uploads/5/6/1/4/56143033/punitive\\_vs\\_restorative\\_school\\_discipline.pdf](https://www.restorativeresources.org/uploads/5/6/1/4/56143033/punitive_vs_restorative_school_discipline.pdf)
- Marcucci, O. (2020). Implicit bias in the era of social desirability: Understanding antiblackness in rehabilitative and punitive school discipline. *Urban Review*, 52(1), 47–74.
- Welsh, R. O., & Little, S. (2018). Caste and control in schools: A systematic review of the pathways, rates and correlates of exclusion due to school discipline. *Children and Youth Services Review*, 94, 315–39.
- Gerlinger, J., Viano, S., Gardella, J. H., Fisher, B. W., Chris Curran, F., & Higgins, E. M. (2021). Exclusionary school discipline and delinquent outcomes: A meta-analysis. *Journal of Youth and Adolescence*, 50, 1493–1509. <https://doi.org/10.1007/s10964-021-01459-3>
- Duarte, C. dP., Salas-Hernández, L., & Griffin, J. S. (2020). Policy determinants of inequitable exposure to the criminal legal system and their health consequences among young people. *American Journal of Public Health*, 110(S1), S43–S49.
- Skiba, R. J., & Losen, D. J. (2016). From reaction to prevention: Turning the page on school discipline. *American Educator*, 39(4), 4.
- Reynolds, C. R., Skiba, R. J., Graham, S., Sheras, P., Conoley, J. C., & Garcia-Vazquez, E. (2008). Are zero tolerance policies effective in the schools?: An evidentiary review and recommendations. *American Psychologist*, 63(9), 852–862.
- Morris, M. (2016). *Pushout: The criminalization of Black girls in schools*. New Press.
- Peterson, J., Densley, J., & Erickson, G. (2021). Presence of armed school officials and fatal and nonfatal gunshot injuries during mass school shootings, United States, 1980–2019. *JAMA Network Open*, 4(2), e2037394.
- Rosenbaum, J. E. (2020). Educational and criminal justice outcomes 12 years after school suspension. *Youth & Society*, 52(4), 515–547.
- Gershoff, E. T., & Font, S. A. (2016). Corporal punishment in U.S. public schools: Prevalence, disparities in use, and status in state and federal policy. *Social Policy Report*, 30, 1.
- Percentage of public schools with security staff present at least once a week, and percentage with security staff routinely carrying a firearm, by selected school characteristics: 2005–06 through 2019–20. (2021). [https://nces.ed.gov/programs/digest/d21/tables/dt21\\_233.70.asp](https://nces.ed.gov/programs/digest/d21/tables/dt21_233.70.asp)
- Whitaker, A., Torres-Guillén, S., Morton, M., Jordan, H., Coyle, S., Mann, A., & Sun, W.-L. (2019). *Cops and no counselors: How the lack of school mental health staff is harming students*. American Civil Liberties Union.
- Muñiz, J. O. (2021). Exclusionary discipline policies, school-police partnerships, surveillance technologies and disproportionality: A review of the school to prison pipeline literature. *The Urban Review*, 53, 735–760. <https://doi.org/10.1007/s11256-021-00595-1>
- Mallett, C. A. (2016). The school-to-prison pipeline: A critical review of the punitive paradigm shift. *Child and Adolescent Social Work Journal*, 33(1), 15–24.
- Skiba, R. J., Mediratta, K., & Rausch, M. K. (2016). *Inequality in school discipline: Research and practice to reduce disparities*. Springer.
- Nowicki, J. M., Doughty, S., Lowe, A. M., Bennett, J., Dye, H., Karty, A., McSweeney, J., Mingus, J., Rebbe, J., Vartivarian, S., Watsula, D., Ayers, J., Bland, D., Carnevale, I., Croake, C., D'Souza, V., Goodwin, G., Hernandez-Saunders, G., Jones, R., ... Steel-Lowney, B. (2018). *K-12 education: Discipline disparities for Black students, boys, and students with disabilities*. United States Government Accountability Office.

22. Bookser, B. A., Ruiz, M., Olu-Odumosu, A., Kim, M., Jarvis, S. N., & Okonofua, J. A. (2021). Context matters for preschool discipline: Effects of distance learning and pandemic fears. *School Psychology*, 36(5), 325–334.
23. Nowicki, J. M. (2020). K-12 education: Education needs to address significant quality issues with its restraint and seclusion data. <https://www.gao.gov/assets/gao-20-345.pdf>
24. Clayback, K. A., & Hemmeter, M. L. (2021). Exclusionary discipline practices in early childhood settings: A survey of child care directors. *Early Childhood Research Quarterly*, 55, 129–136.
25. Verdugo, R. R. (2002). Race-ethnicity, social class, and zero-tolerance policies: The cultural and structural wars. *Education and Urban Society*, 35(1), 50–75.
26. Gilmore, R. W. (2008). *1. Forgotten places and the seeds of grassroots planning. Engaging contradictions*. University of California Press.
27. Simon, J. (2007). *Governing through crime: How the war on crime transformed American democracy and created a culture of fear*. Oxford University Press.
28. Morris, M. W. (2016). *Pushout: The criminalization of Black girls in schools*. New Press.
29. Rudolph, L., Caplan, J., Ben-Moshe, K., & Dillon, L. (2013). *Health in all policies: A guide for state and local governments*. American Public Health Association and Public Health Institute.
30. Building a Culture of Health. (2020). <https://www.rwjf.org/en/how-we-work/building-a-culture-of-health.html>
31. Neely, A. N., Ivey, A. S., Duarte, C., Poe, J., & Irsheid, S. (2020). Building the transdisciplinary resistance collective for research and policy: Implications for dismantling structural racism as a determinant of health inequity. *Ethnicity & Disease*, 30(3), 381–388.
32. Duarte, C., Cameron, D. B., Kwan, A. T., Bertozzi, S. M., Williams, B. A., & McCoy, S. I. (2022). COVID-19 outbreak in a state prison: A case study on the implementation of key public health recommendations for containment and prevention. *BMC Public Health*, 22(1), 977.
33. Robertson, T., Benzeval, M., Whitley, E., & Popham, F. (2015). The role of material, psychosocial and behavioral factors in mediating the association between socioeconomic position and allostatic load (measured by cardiovascular, metabolic and inflammatory markers). *Brain, Behavior, and Immunity*, 45, 41–49.
34. World Health Organization. (2010). Conceptual framework of the social determinants of health. *A conceptual framework for action on the social determinants of health: Social determinants of health discussion paper 2 (policy and practice)*.
35. Williams, D. R. (1997). Race and health: Basic questions, emerging directions. *Annals of Epidemiology*, 7(5), 322–333.
36. Krieger, N. (2001). Theories for social epidemiology in the 21st century: An ecosocial perspective. *International Journal of Epidemiology*, 30(4), 668–677.
37. Ford, C. L., & Airhihenbuwa, C. O. (2010). Critical race theory, race equity, and public health: Toward antiracism praxis. *American Journal of Public Health*, 100(1), S30–S35.
38. Gee, G. C., Hing, A., Mohammed, S., Tabor, D. C., & Williams, D. R. (2019). Racism and the life course: Taking time seriously. *American Journal of Public Health*, 109(S1), S43–S47.
39. Heitzeg, N. A. (2015). 'Whiteness,' criminality, and the double standards of deviance/social control. *Contemporary Justice Review*, 18(2), 197–214.
40. Pfohl, S. J. (1994). *Images of deviance and social control: A sociological history*. McGraw-Hill.
41. Bonilla-Silva, E. (2006). *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States*. Rowman & Littlefield Publishers.
42. Travis, J., Western, B., & Redburn, F. S. (2014). The growth of incarceration in the United States: Exploring causes and consequences.
43. Bonilla-Silva, E. (2015). The structure of racism in color-blind, "post-racial" America. *American Behavioral Scientist*, 59(11), 1358–76.
44. Bailey, Z. D., Feldman, J. M., & Bassett, M. T. (2020). How structural racism works — Racist policies as a root cause of U.S. racial health inequities. *New England Journal of Medicine*, 384(8), 768–773.
45. Kuperminc, G. P., Leadbeater, B. J., & Blatt, S. J. (2001). School social climate and individual differences in vulnerability to psychopathology among middle school students. *Journal of School Psychology*, 39(2), 141–159.
46. Zajacova, A., & Lawrence, E. M. (2018). The relationship between education and health: Reducing disparities through a contextual approach. *Annual Review of Public Health*, 39(1), 273–289.
47. Markovitz, J. H., Matthews, K. A., Whooley, M., Lewis, C. E., & Greenlund, K. J. (2004). Increases in job strain are associated with incident hypertension in the CARDIA Study. *Annals of Behavioral Medicine*, 28(1), 4–9.
48. Lynch, S. M. (2006). Explaining life course and cohort variation in the relationship between education and health: The role of income. *Journal of Health and Social Behavior*, 47(4), 324–338.
49. Lunau, T., Siegrist, J., Dragano, N., & Wahrendorf, M. (2015). The association between education and work stress: Does the policy context matter? *PLoS One*, 10(3), e0121573.
50. Petteway, R., Mujahid, M., & Allen, A. (2019). Understanding embodiment in place-health research: Approaches, limitations, and opportunities. *Journal of Urban Health*, 96(2), 289–299.
51. Staufenbiel, S. M., Penninx, B. W. J. H., Spijker, A. T., Elzinga, B. M., & van Rossum, E. F. C. (2013). Hair cortisol, stress exposure, and mental health in humans: A systematic review. *Psychoneuroendocrinology*, 38(8), 1220–1235.
52. Hong, J. S., Davis, J. P., Sterzing, P. R., Yoon, J., Choi, S., & Smith, D. C. (2014). A conceptual framework for understanding the association between school bullying victimization and substance misuse. *American Journal of Orthopsychiatry*, 84(6), 696–710.
53. Bucchianeri, M. M., Eisenberg, M. E., Wall, M. M., Piran, N., & Neumark-Sztainer, D. (2014). Multiple types of harassment: Associations with emotional well-being and unhealthy behaviors in adolescents. *Journal of Adolescent Health*, 54(6), 724–729.
54. Monahan, K. C., VanDerhei, S., Bechtold, J., & Cauffman, E. (2014). From the school yard to the squad car: School discipline, truancy, and arrest. *Journal of Youth and Adolescence*, 43(7), 1110–1122.
55. Simons, R. L., Lei, M.-K., Klopach, E., Beach, S. R. H., Gibbons, F. X., & Philibert, R. A. (2021). The effects of social adversity, discrimination, and health risk behaviors on the accelerated aging of African Americans: Further support for the weathering hypothesis. *Social Science & Medicine*, 282, 113169. <https://doi.org/10.1016/j.socscimed.2020.113169>
56. Michaels, E. K., Reeves, A. N., Thomas, M. D., Price, M. M., Hasson, R. E., Chae, D. H., & Allen, A. M. (2019). Everyday racial discrimination and hypertension among midlife african american women: Disentangling the role of active coping dispositions versus active coping behaviors. *International Journal of Environmental Research and Public Health*, 16(23), 4759. <https://doi.org/10.3390/ijerph16234759>
57. Whear, R., Marlow, R., Boddy, K., Ukoumunne, O. C., Parker, C., Ford, T., Thompson-Coon, J., & Stein, K. (2014). Psychiatric disorder or impairing psychology in children who have been excluded from school: A systematic review. *School Psychology International*, 35(5), 530–543.
58. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, 6(7), e1000100.
59. Veritas Health Innovation. (2016). *Covidence systematic review software*. Veritas Health Innovation.
60. Racz, S. J., Saha, S., Trent, M., Adger, H., Bradshaw, C. P., Goldweber, A., & Cauffman, E. (2016). Polysubstance use among minority



- adolescent males incarcerated for serious offenses. *Child & Youth Care Forum*, 45(2), 205–220.
61. Ramey, D. M. (2016). The influence of early school punishment and therapy/medication on social control experiences during young adulthood. *Criminology*, 54, 113–141.
  62. Elster, A. B., Ketterlinus, R., & Lamb, M. E. (1990). Association between parenthood and problem behavior in a national sample of adolescents. *Pediatrics*, 85(6), 1044–1050.
  63. Gillmore, M. R., Butler, S. S., Lohr, M. J., & Gilchrist, L. (1992). Substance use and other factors associated with risky sexual behavior among pregnant adolescents. *Family Planning Perspectives*, 24(6), 255–261.
  64. Bottiani, J. H., Bradshaw, C. P., & Mendelson, T. (2017). A multilevel examination of racial disparities in high school discipline: Black and White adolescents' perceived equity, school belonging, and adjustment problems. *Journal of Educational Psychology*, 109, 532–545.
  65. Lee, G. P., Martins, S. S., Pas, E. T., & Bradshaw, C. P. (2014). Examining potential school contextual influences on gambling among high school youth. *American Journal on Addictions*, 23(5), 510–517.
  66. Paek, H. J., Hove, T., & Oh, H. J. (2013). Multilevel analysis of the impact of school-level tobacco policies on adolescent smoking: The case of Michigan. *Journal of School Health*, 83(10), 679–689.
  67. Fothergill, K. E., Ensminger, M. E., Green, K. M., Crum, R. M., Robertson, J., & Juon, H. S. (2008). The impact of early school behavior and educational achievement on adult drug use disorders: A prospective study. *Drug and Alcohol Dependence*, 92(1-3), 191–199.
  68. Gould, M. S., Fisher, P., Parides, M., Flory, M., & Shaffer, D. (1996). Psychosocial risk factors of child and adolescent completed suicide. *Archives of General Psychiatry*, 53(12), 1155–1162.
  69. Rosenbaum, J. E. (2020). School suspension predicts trichomoniasis five years later in a matched sample. *BMC Public Health*, 20(1), 88.
  70. Murray, J., Farrington, D. P., & Sekol, I. (2012). Children's antisocial behavior, mental health, drug use, and educational performance after parental incarceration: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 175–210.
  71. Moitra, M., Santomauro, D., Degenhardt, L., Collins, P. Y., Whiteford, H., Vos, T., & Ferrari, A. (2021). Estimating the risk of suicide associated with mental disorders: A systematic review and meta-regression analysis. *Journal of Psychiatric Research*, 137, 242–249.
  72. Onyeka, I. N., Maguire, A., Ross, E., & O'Reilly, D. (2020). Does physical ill-health increase the risk of suicide? A census-based follow-up study of over 1 million people. *Epidemiology and Psychiatric Sciences*, 29, e140.
  73. Conrad, P., & Bergey, M. (2015). Medicalization: Sociological and anthropological perspectives. In J. D. Wright (Ed.), *International Encyclopedia of the social & behavioral sciences* (pp. 105–109, 2nd ed.). Elsevier.
  74. Rushton, J. L., Forcier, M., & Schectman, R. M. (2002). Epidemiology of depressive symptoms in the National Longitudinal Study of Adolescent Health. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(2), 199–205.
  75. Cottrell, C. (2018). Racial differences in anger and depression as mediators in the relationship between suspension and juvenile delinquency: A test of general strain theory. *Journal for the Advancement of Educational Research International*, 12, 58–69.
  76. Kramer, U., Temes, C. M., Magni, L. R., Fitzmaurice, G. M., Aguirre, B. A., Goodman, M., & Zanarini, M. C. (2017). Psychosocial functioning in adolescents with and without borderline personality disorder. *Personality and Mental Health*, 11(3), 164–170.
  77. Hemphill, S. A., Toumbourou, J. W., Herrenkohl, T. I., McMorris, B. J., & Catalano, R. F. (2006). The effect of school suspensions and arrests on subsequent adolescent antisocial behavior in Australia and the United States. *Journal of Adolescent Health*, 39, 736–744.
  78. Talluri, R., Wilkinson, A. V., Spitz, M. R., & Shete, S. (2014). A risk prediction model for smoking experimentation in Mexican American youth. *Cancer Epidemiology, Biomarkers & Prevention*, 23(10), 2165–2174.
  79. Hemphill, S. A., Heerde, J. A., Herrenkohl, T. I., Toumbourou, J. W., & Catalano, R. F. (2012). The impact of school suspension on student tobacco use: A longitudinal study in Victoria, Australia, and Washington State, United States. *Health Education & Behavior*, 39(1), 45–56.
  80. Alexander, C. S., Somerfield, M. R., Ensminger, M. E., Kim, Y. J., & Johnson, K. E. (1995). Gender differences in injuries among rural youth. *Injury Prevention*, 1(1), 15–20.
  81. Fowler, M. G., Davenport, M. G., & Garg, R. (1992). School functioning of US children with asthma. *Pediatrics*, 90(6), 939–944.
  82. Slade, E. P. (2004). Racial/ethnic disparities in parent perception of child need for mental health care following school disciplinary events. *Mental Health Services Research*, 6(2), 75–92.
  83. Khan, S., Pinckney, R. G., Keeney, D., Frankowski, B., & Carney, J. K. (2011). Prevalence of food insecurity and utilization of food assistance program: An exploratory survey of a Vermont middle school. *Journal of School Health*, 81(1), 15–20.
  84. Gennetian, L. A., Seshadri, R., Hess, N. D., Winn, A. N., & Goerge, R. M. (2016). Supplemental nutrition assistance program (SNAP) benefit cycles and student disciplinary infractions. *Social Service Review*, 90(3), 403–433.
  85. Bond, T. N., Carr, J. B., Packham, A., & Smith, J. (2021). *Hungry for success? SNAP timing, high-stakes exam performance, and college attendance*. National Bureau of Economic Research.
  86. Shankar, P., Chung, R., & Frank, D. A. (2017). Association of food insecurity with children's behavioral, emotional, and academic outcomes: A systematic review. *Journal of Developmental and Behavioral Pediatrics*, 38(2), 135–150.
  87. Sinha, R. (2008). Chronic stress, drug use, and vulnerability to addiction. *Annals of the New York Academy of Sciences*, 1141, 105–130.
  88. Babey, S., Wolstein, J., Becker, T., & Scheitler, A. (2019). *School discipline practices associated with adolescent school connectedness and engagement*. UCLA Center for Policy Research.
  89. Farkas, K., Duarte, C. dP., & Ahern, J. (2022). Injuries to children and adolescents by law enforcement: An analysis of California emergency department visits and hospitalizations, 2005–2017. *JAMA Pediatrics*, 176(1), 89–91. <https://doi.org/10.1001/jamapediatrics.2021.2939>
  90. Smith, K. E., & Pollak, S. D. (2020). Early life stress and development: Potential mechanisms for adverse outcomes. *Journal of Neurodevelopmental Disorders*, 12(1), 34.
  91. Sheth, C., McGlade, E., & Yurgelun-Todd, D. (2017). Chronic stress in adolescents and its neurobiological and psychopathological consequences: An RDoC perspective. *Chronic Stress (Thousand Oaks)*, 1, 2470547017715645. <https://doi.org/10.1177/2470547017715645>
  92. van Heeringen, C. (2001). Suicide in adolescents. *International Clinical Psychopharmacology*, 16(2), S1–S6.
  93. Hawton, K., Kingsbury, S., Steinhardt, K., James, A., & Fagg, J. (1999). Repetition of deliberate self-harm by adolescents: The role of psychological factors. *Journal of Adolescence*, 22(3), 369–378.
  94. Steinhoff, A., Ribeaud, D., Kupferschmid, S., Raible-Destan, N., Quednow, B. B., Hepp, U., Eisner, M., & Shanahan, L. (2021). Self-injury from early adolescence to early adulthood: Age-related course, recurrence, and services use in males and females from the community. *European Child & Adolescent Psychiatry*, 30(6), 937–951.
  95. Farkas, K., Matthay, E. C., Rudolph, K. E., Goin, D. E., & Ahern, J. (2019). Mental and substance use disorders among legal intervention injury cases in California, 2005–2014. *Preventive Medicine*, 121, 136–140.
  96. Mathers, M., Toumbourou, J. W., Catalano, R. F., Williams, J., & Patton, G. C. (2006). Consequences of youth tobacco use: A review of prospective behavioural studies. *Addiction*, 101(7), 948–958.

97. Tobin, T. J., & Sugai, G. M. (1999). Using sixth-grade school records to predict school violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders*, 7(1), 40–53.
98. Krieger, N. (2014). Discrimination and health inequities. In L. F. Berkman, I. Kawachi, & M. M. Glymour (Eds.), *Social epidemiology* (pp. 63–125). Oxford University Press.
99. Ramey, D. M. (2018). The social construction of child social control via criminalization and medicalization: Why race matters. *Sociological Forum*, 33(1), 139–64.
100. Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). *Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions*. Yale University Child Study Center.
101. Ramsey, C. M., Spira, A. P., Parisi, J. M., & Rebok, G. W. (2016). School climate: Perceptual differences between students, parents, and school staff. *School Effectiveness and School Improvement*, 27(4), 629–41.
102. Victora, C. G., Huttly, S. R., Fuchs, S. C., & Olinto, M. T. (1997). The role of conceptual frameworks in epidemiological analysis: A hierarchical approach. *International Journal of Epidemiology*, 26(1), 224–7.
103. Bacher-Hicks, A., Billings, S., & Deming, D. (2021). Proving the School-to-Prison Pipeline: Stricter middle schools raise the risk of adult arrest. *Education Next*, 21(4), 52–7.
104. Annamma, S. A., Anyon, Y., Joseph, N. M., Farrar, J., Greer, E., Downing, B., & Simmons, J. (2019). Black girls and school discipline: The complexities of being overrepresented and understudied. *Urban Education*, 54(2), 211–42.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Duarte, C. dP., Moses, C., Brown, M., Kajeepeta, S., Prins, S. J., Scott, J., & Mujahid, M. S. (2022). Punitive school discipline as a mechanism of structural marginalization with implications for health inequity: A systematic review of quantitative studies in the health and social sciences literature. *Ann NY Acad Sci.*, 1–24.  
<https://doi.org/10.1111/nyas.14922>