The Hidden Health Care Crisis Behind Bars: A Randomized Trial to Accredit U.S. Jails

Marcella Alsan*
Harvard Kennedy School
NBER

Crystal S. Yang
Harvard Law School
NBER

March 18, 2025

Abstract

The U.S. has one of the highest incarceration rates in the world, with over seven million admissions to jails each year. Incarcerated individuals are the only group in the U.S. that have a constitutional right to receiving "reasonably adequate" health care. Yet, there is little oversight and funding for health care in jails, where illness and mortality are rampant. In this study, we randomize the offer of health care accreditation to 44 jails across the U.S. Surveys of staff indicate that accreditation improves coordination between health and custody staff. We also find that accreditation improves quality standards and reduces mortality among the incarcerated, which is three times higher among control facilities than official estimates suggest. These health gains are realized alongside suggestive reductions in six-month recidivism, such that accreditation is highly cost effective.

Keywords: health, accreditation, mortality, incarcerated, jail, crime, recidivism
JEL Codes: K10, I13, I14, I18

*Alsan: Kennedy School of Government, Harvard University, 79 John F. Kennedy St., Rubenstein Bldg R403, Cambridge, MA 02138 (email: marcella_alsan@hks.harvard.edu). Yang: Harvard Law School, 1585 Massachusetts Ave., Hauser 516, Cambridge, MA 02138 (email: cyang@law.harvard.edu). We gratefully acknowledge funding from JPAL North America, the William T. Grant Foundation, the Preparedness and Treatment Equity Coalition, Arnold Ventures, and the Malcolm Wiener Center at the Harvard Kennedy School. We thank seminar participants at NBER Law and Economics, NBER Crime, Harvard Law School, Harvard Kennedy School, Vancouver School of Economics, UCLA Anderson, MIT, Minnesota Department of Management and Budget, and Dartmouth. We thank the National Commission on Correctional Health Care (NCCHC), especially Chief Executive Officer Deborah Ross and Vice-President of Accreditation Amy Panagopoulos. We also thank jail custody and health staff, incarcerated individuals, and county sheriffs for their time and participation. We thank Jim Martin, Dr. Cara Buchanan Muñoz, George Hutchins, and Dr. Nicholos Joseph for sharing their expertise. This work would not have been possible without the effort of our dedicated project coordinators: Tasfia Jahangir and Nadia Huffman. Miguel Purroy, Qing Liu, Shipra Karan, Xingyou Ye, Vanessa Jia, Nada Shalash, Nick Shankar, and Lukas Leister provided excellent research assistance. Finally, we thank the numerous college, law, public policy, and medical students who worked on the project: James Jolin, Lucy Tu, Gwen Byrne, Aisha Oshilaja, Pierre Jolin, Josh Alpert, Daniel Tochimani-Hernandez, Farrah Kalach, Janine Robredo, Sophia Fend, Bolin Shen, Yedith Ortiz, Emily Shah, Susan Takang, Kate De Groote, Worthy Cho, Israa Alzami, Anne Deng, Christine Smith, Aydin Alsan, Achele Agada, Cana Nakase, Sarah Kliment, Kathryn Li, Kaitlyn Ham, Tatheer Adnan, Anita Onyimah, Folukemi Olufidipe, Meisui Liu, Phoebe Canagarajah, Lea Han, Allie Oh, Louis Lin, Jennifer Ding, Zach Channen, Nik Zappia, Fiona Doyle, Christian Carson, Drew Lane, Jack Haney and Haley Dougherty. Angelique Aquatella and Karine Van Der Straeten provided helpful feedback. The experiment was registered on the AEA RCT registry (RCT ID: AEARCTR-0007516).

I Introduction

Serving as the point of entry into the criminal justice system, U.S. jails have over seven million admissions per year, with over 600,000 individuals held in jails on any given day, the vast majority of whom are unconvicted (Zeng 2023). Jails are typically considered short-term detention facilities, but mortality in jails has increased in the last decade, often occurring within hours or days of admission (Adler and Chen 2023). Drug- or alcohol-related deaths in jail have increased by a factor of four over the same time period, and the official suicide death rate remains at two times the national average (Wang 2021; Carson 2021). Our own data reveal a mortality problem that is much more severe, with death rates *three times higher* than those reported to federal authorities, reflecting longstanding difficulties with accurate reporting and transparency of deaths in custody (Mitchell Jr and Aronson 2023).[1] Thus, improving the quality of health care provided in jails is of utmost policy importance.

One commonly used method to ascertain quality in settings of information frictions is accreditation. Broadly defined, accreditation is a process during which an agency assesses a product's compliance with industry quality standards. However, whether and to what extent accreditation is effective is unsettled: models of accreditation often deliver multiple equilibria depending on whether a certifying agency behaves opportunistically – colluding or shirking – or as an honest broker (Strausz 2005). As a practical matter, accreditation is widespread in the U.S. health care system. Typically, the process involves the development of standards and an independent, on-site audit of the health care provider (Sechrest 1976). In the U.S. civilian health care system, for example, such accreditation is typically done by the Joint Commission (TJC), and attaining "deemed status" through the TJC allows health care facilities to charge Medicare and Medicaid for reimbursement. As a result, nearly all U.S. civilian patients receive care from TJC-accredited hospitals (The Joint Commission 2024).

This is not true for incarcerated patients, however, as only a small minority of correctional facilities have sought voluntary accreditation by a third-party agency (Rich, Allen and Williams 2015). This relatively low rate is due in part to the lack of financial incentives, as medical care during incarceration is not covered by federal funds.[2] Yet the need for quality care is critical for the incarcerated population, who not only suffer from a dual burden of communicable and non-communicable disease, but are also the only U.S. population constitutionally guaranteed a right to adequate health care since the landmark Supreme Court decision in *Estelle v. Gamble* (*1976*) (Alsan et al. 2023). Correctional facilities are thus legally required to provide adequate care, but are led by individuals with limited medical knowledge who often contract with for-profit vendors in the face of rising medical costs, further heightening concerns about access and quality (Bedard and Frech III 2009; Henrichson, Rinaldi and Delaney 2015; Berwick, Beckman and Gondi 2021). Correctional leaders also face a principal with multiple-agents problem as they depend on custody and medical staff, two agents whose cooperation is essential to ensure delivery of health care (Alchian and Demsetz 1972; Holmstrom 1982).

In this study, we evaluate whether accreditation affects the quality of care, mortality, and staff outcomes in U.S. jails. In theory, the effects of accreditation are ambiguous. If standards focus on items that do not improve health or if they serve merely as a "rubber stamp," accreditation could have an inconclusive or neg-

[1] Reuters data on deaths in custody also show a much higher mortality rate than the Bureau of Justice Statistics (BJS) data collected by the Department of Justice (DOJ).

[2] The Social Security Act (Sec. 1905(a)(A)) prohibits use of federal funds and services...for medical care provided to inmates of a public institution (Social Security Amendments 1965; Edmonds 2020).

ative effect on health care quality and outcomes by diverting attention and resources to less useful activities or by fostering complacency with poor care.[3] If, on the other hand, accreditation provides information to improve coordination between agents and allows the principal to observe individual agent effort, it may have a positive and measurable effect on health outcomes.

To estimate the causal effect of accreditation, we conducted a randomized controlled trial with 44 jails across the U.S. spanning four years. The participating jails are representative of small to medium-sized jails in the country. Stratifying jails by size and enrollment timing, we randomized jails to undergo an accreditation process administered by the National Commission on Correctional Health Care (NCCHC), a non-profit correctional health care accrediting agency founded in 1983 (National Commission on Correctional Health Care 2018; Gibson and Phillips 2016). NCCHC accreditation is based on quality standards that comprise seven core categories ranging from governance and administration to patient care and treatment. To measure compliance, NCCHC performs on-site audits and issues a report, after which a committee of experts blind to facility identity decides on the accreditation status. NCCHC's accreditation process is regarded as a "gold standard" among many practitioners and policymakers and obtaining NCCHC accreditation is often regarded as a signal of constitutionally adequate health care by courts, although critics allege it is ineffective.

In our study, 22 treatment facilities received a generous subsidy towards the accreditation fee and 22 control facilities were offered a more modest subsidy to commence accreditation at the end of the study period approximately 24 months after enrollment. We developed and administered several novel survey instruments at both baseline and endline, including questionnaires for facility leadership, custody and medical staff, and independent audits of medical records and death logs. At endline, we conducted qualitative interviews with jail leadership and incarcerated individuals. These survey instruments were created and coded by dozens of undergraduate, medical, and law students to ensure expertise in interpreting clinical and legal data. Lastly, we conducted a survey with experts comprising jail administrators and health care providers, policy experts and academics, and the formerly incarcerated. Respondents were asked to weight the importance of different standards and predict the outcome of our experiment and heterogeneity.

Using these survey instruments, we find that accreditation has statistically significant and economically meaningful effects on a range of outcomes. At the end of our study, 13 of 22 treatment facilities will have successfully received NCCHC accreditation.[4] Instrumenting with treatment assignment for completion of the accreditation process in a two-stage-least-squares specification, we find that completing the accreditation process improves measures of horizontal coordination (between custody and medical staff) and staff preparation, with null effects for staff perceptions of respect for incarcerated individuals. In addition, accreditation improves compliance with meta-indices of quality standards, driven by improvements on standards governing "personnel training" and "patient care and treatment." Notably, outcomes such as additional staff and large capital investments do not change with accreditation, as county budgets are relatively constrained and standards focus mainly on improving processes and procedures.

In terms of health outcomes, we find that completion of the NCCHC accreditation process substantially decreases mortality as measured by average deaths per month, with a reduction of about 90% in the half-year following the expected time of the NCCHC on-site visit. Our mortality effects are considerable and robust to a range of different estimators (*e.g.*, non-linear) samples and specifications. Specifically, we

[3] With respect to an investigation into the accreditation process conducted by the American Correctional Association (ACA), Senator Elizabeth Warren reported that "it appears the ACA is a conflicted party with twisted incentives, a lack of transparency, and lax inspection policies that appear to have turned accreditation into a rubber-stamp process that does little to hold facilities accountable."

[4] One of these thirteen is in the final stages of finalizing its status.

find similar effects using reduced form estimates or the *fraction of process completed* as measured by either NCCHC or by the research team at Harvard as the endogenous variable. This latter specification admits the possibility that completing *some* of the process has salubrious effects. In practice, the point estimates are indistinguishable as most who exited the process did so without making any improvements, but the standard errors are smaller using a process versus completion measure. We also confirm that our mortality effects are unlikely to be driven by reporting bias, as our findings can be replicated using only media reports of deaths. Rather, it seems the magnitude of our effects are due to the fact that there is considerable scope for improvement among U.S. jails, where information frictions and coordination problems are severe and where health outcomes are much worse than widely believed. Such an interpretation is supported by our in-depth qualitative work, which shows that the accreditation process helped align custody and medical staff, and provided important information on how to process and screen inmates, particularly in the very first days of incarceration where mortality tends to be the highest.

From a welfare perspective, one question is whether our intervention is simply displacing deaths until after release and not actually saving lives. Although true on some fundamental level, as all any treatment can do is delay death, a significant portion of the deaths in custody are attributable to the conflation of crisis and confinement – for example, suicidality exacerbated by segregation; drug use exacerbated by intoxication or withdrawal; or pneumonia progressing to sepsis exacerbated by overcrowding. Thus, jails function similarly to emergency departments, admitting relatively young and able-bodied individuals who are acutely in crisis and often ill. As a result, how a jail screens, processes, and triages incarcerated individuals in the initial hours and days has important ramifications for survival.

To assess heterogeneity, we rely on the predictions from our expert survey. The three factors selected by experts as most likely to affect the impact of NCCHC accreditation were the number of medical staff (FTE), the health care quality, and the quality of jail management. Yet, we find modest evidence that any factor selected by experts predicts heterogeneity. We also find suggestive evidence that accreditation reduces the probability of six-month recidivism but there is no detectable effect on litigation.

Broadly, our study shows that accreditation can improve health care and save lives in U.S. jails. These findings contribute to a growing literature in both economics and public health studying health care and health conditions in U.S. jails.[5] For example, the public health literature generally reports harmful effects of jails, including higher rates of overdose deaths immediately upon release (Binswanger et al. 2007) and higher rates of deaths due to carceral exposure over the life course (NASEM 2019; Daza, Palloni and Jones 2020). More recently, Norris, Pecenco and Weaver (2024), using an event study framework, find detrimental effects of release from incarceration in Ohio on mortality. The authors interpret this as evidence of high mortality risk among community-dwelling men that is lowered by the incapacitation effects of incarceration. Similar effects on prison improving health outcomes among some groups of inmates were found in the context of Scandinavian countries, where healthcare is described by the authors as "high quality" and time in prison increases interaction with psychologists, doctors, and addiction treatment (Hjalmarsson and Lindquist 2022; Bhuller, Khoury and Løken n.d.).

Economists have also explored the link between conditions of confinement and crime, with studies by Katz, Levitt and Shustorovich (2003) and Bedard and Helland (2004) finding that higher mortality rate in prison and prison punitiveness as measured by distance from residence reduces crime due to a deterrence

[5]See Batistich, Evans and Phillips (2021); Patterson (2010); Spaulding et al. (2011); Wildeman et al. (2016); Norris, Pecenco and Weaver (2024); Binswanger et al. (2007); Smith (2013); Bovell-Ammon et al. (2021); Ruch et al. (2021)

effect. On the other hand, since over 95% of those incarcerated in jail return to their communities, providing services while incarcerated could improve community health and reduce the cycle of violence, with growing evidence of the effectiveness of rehabilitative programming (Doleac 2023; Alsan et al. 2024). For instance, improving health care in jails may mitigate community spread of HIV/AIDS and COVID-19 and prevent the widening of disparities across communities (Johnson and Raphael 2009; Reinhart and Chen 2021).

Our paper also contributes to an important theoretical and empirical literature on quality improvements in healthcare, and the impact of quality assurance mechanisms (Bloom et al. 2015, 2020; Dranove and Jin 2010). Theoretically, the effect of quality disclosure is ambiguous and the limited empirical evidence is mixed, but some studies find that quality disclosure can improve welfare. Jin and Leslie (2003), for example, find that restaurant hygiene grade cards reduced hospitalizations from food-borne diseases. Chen (2008) studies the effect of the Nursing Home Quality Initiative, finding that lower quality nursing homes improved relative to high quality homes, with larger improvements in more competitive markets. In the education context, studies have also found that teacher accreditation via the National Board for Professional Teaching Standards (NBPTS) can identify more effective teachers but yields mixed effects on changing teachers' performance (Goldhaber and Anthony 2007; Kane, Rockoff and Staiger 2008; Clotfelter, Ladd and Vigdor 2010). We contribute to this literature by studying the effects of quality assurance via accreditation for health care in jails.

Finally, we build on a literature identifying effects of health care accreditation specifically. A review by TJC found that 80% of the 189 published research studies on accreditation showed a positive impact on quality, although all are observational in nature and were not conducted within correctional settings (The Joint Commission 2021), and a recent systemic review of 21 higher quality studies on hospital accreditation noted that findings were inconsistent (Alhawajreh, Paterson and Jackson 2023). Another study evaluating Denmark's mandatory national accreditation program found that it improved quality of care overall, particularly in low-performing areas, with gains leveling off after attainment (Bogh et al. 2017). Additionally, a study which randomized accreditation of 20 hospitals in South Africa found that accreditation improved quality metrics (Shaw 2003). We contribute to this existing work by conducting, to the best of our knowledge, the first randomized controlled trial across U.S. jails, and the first randomized trial of accreditation in any medical context in the U.S.

The remainder of the paper is organized as follows. Section II provides legal and medical background. Section III presents a conceptual framework of accreditation. Section IV provides details on intervention and experimental design and Section V describes the data we used. Section VI describes our findings, Section VII describes heterogeneity analyses based on an original expert survey, and Section VIII provides qualitative evidence. Section IX concludes.

II Institutional Background

II.1 Medical Context

The U.S. incarcerates more individuals per capita than many other countries in the world (Fair and Walmsley 2024). In 2022, U.S. jails had 7.3 million admissions (compared to approximately 34 million hospital admissions in the same year (Health Forum 2024)), with over 600,000 individuals held in jails on any given day (Zeng 2023). Jails represent a critical juncture because they are the point of entry for the criminal justice

system and where many individuals are first screened and triaged upon booking.

Individuals who enter the jail system suffer from a dual burden of chronic and infectious disease (Wang 2021). The incarcerated population generally comes from lower-income households, with substantially greater physical and mental health needs than the general population; they have higher rates of tuberculosis, HIV, hepatitis, diabetes, and psychiatric illness (Maner et al. 2022; Curran et al. 2023). More than half of incarcerated individuals have a mental health problem, a substance use disorder, or both, and they experience higher rates of geriatric conditions than the age-matched general population (Puglisi and Wang 2021; Greene et al. 2018).

As a result, the occurrence of adverse events is commonplace and rising in jails. For example, mortality rates have increased by 11% from 2000 to 2019, with a *reported* average mortality rate in U.S. jails of 1.5 deaths per 1,000 and a mortality rate of almost 2 deaths per 1,000 among the unconvicted jail population in 2019, approximately half the mortality rate for the adjusted general population (Carson 2021; Adler and Chen 2023). During the same time period, drug- or alcohol-related deaths in jail increased by a factor of four, and the suicide death rate remained high at two times the national average (Wang 2021; Carson 2021).

Accurate reporting on deaths in jails, however, is problematic. Despite reauthorization of the Deaths in Custody Reporting Act (DCRA) in 2013, disclosure has been seriously flawed for several reasons (McCann 2024; Congressional Research Service 2023). First, the penalty for noncompliance (up to a 10% reduction in the Edward Byrne Memorial Justice Association Grant - JAG) is relatively weak and enforced at the state not local level. Second, the Department of Justice (DOJ) moved responsibility for collecting DCRA data from the Bureau of Justice Statistics to its Bureau of Justice Assistance, which is less well-equipped for such activity, citing concerns a that a federal statistical agency could not be involved with policymaking. Third, the DOJ has been slow to release data, yet even when it does there are serious flaws: three federal agencies including the General Accounting Office (GAO) found substantial underreporting of deaths-in-custody to the Bureau of Justice Assistance.[6] Indeed, among our own data, the death rate based on death logs (4.5 per 1,000) and Reuters-reported deaths is substantially higher than official statistics published by the DOJ.

II.2 Legal Context

Constitutional Standards and Relevant Statutes: The legal community, in conjunction with the medical community, has played an important role in shaping the delivery of health care in U.S. jails through the development of constitutional and statutory requirements.[7] In the 1976 landmark decision *Estelle v. Gamble*, the Supreme Court held that deliberate indifference to the serious medical needs of incarcerated people violates the Eighth Amendment's prohibition against cruel and unusual punishment (American Medical Association 1979; Greifinger 2007; Alsan et al. 2023). In *Estelle*, the Court construed the Eighth Amendment as embodying "broad and idealistic concepts of dignity, civilized standards, humanity, and decency," leading the Court to conclude that the "deliberate indifference to the serious medical needs of prisoners constitute[d] the 'unnecessary and wanton infliction of pain' prohibited by the amendment."[8] In *Helling v. McKinney* (1993), the Court expanded on the *Estelle* decision, finding that the Eighth Amendment also "protect[ed] against future harm to inmates," laying the groundwork for prevention in addition to treatment for existing conditions. Moreover, the Court in *West v. Atkins* (1988) extended liability to privately contracted medical

[6] The GAO report found 1,000 unreported deaths in 2021 compared to other sources and noted 70% of records submitted to the Department of Justice were missing key information (GAO 2022).

[7] See Appendix Section L.2 for further details.

[8] *Estelle v. Gamble*, 429 U.S. 97 (1976).

staff, holding that private physicians contracted to work at state prisons act under the color of state law, and thus can be sued for providing inadequate medical treatment (Pew Charitable Trusts 2014).

While this constitutional right exists, the *Estelle* standard for establishing an Eighth Amendment violation is a very high bar to meet. For instance, federal courts interpret deliberate indifference to mean treatment that "must be so grossly incompetent, inadequate, or excessive as to shock the conscience or to be intolerable to fundamental fairness."[9] Negligent medical treatment and even medical malpractice does not meet the deliberate indifference requirement. In addition, incarcerated individuals face numerous barriers to filing suit in the first place, ranging from statutory hurdles imposed by the Prison Litigation Reform Act to difficulties with obtaining counsel (Alsan et al. 2023).

Statutory barriers to the financing of health care in correctional facilities further hamper the ability of jails to fulfill these constitutional obligations. Specifically, the 1965 Medicaid Inmate Exclusion Policy prohibits Medicaid from reimbursing most health care delivered to the incarcerated (Edmonds 2020).[10] As a result, the cost of medical care in corrections continues to rise, with corrections representing the largest expenditure for state budgets after Medicaid and with one-sixth of prison budgets dedicated to health care (Pew Charitable Trusts 2014). This trend is expected to continue given medical inflation, aging of the incarcerated population and increasing mental health disorders (Williams, DiTomas and Pachynski 2021). Aiming to contain costs and limit liability, sheriffs and jails officials with limited medical knowledge increasingly contract with private for-profit vendors to provide health care whose quality is difficult to observe.[11] In addition to these information frictions, there are coordination problems since both custody and medical staff must work together to ensure delivery of health care as a team despite having different objectives. In such a circumstance: what, if any, is the role of health care accreditation?

II.3 Accreditation

Accreditation in health care: Accreditation is a process designed to ensure quality of organizations in settings of limited information.[12] Typically, the accreditation process involves the development of standards and an independent, on-site audit by a third-party (Sechrest 1976). The theoretical effects of accreditation on quality and associated outcomes are ambiguous and depend on whether the third-party is an honest broker (Dranove and Jin 2010). Empirical findings on the effects of accreditation on quality in health care, where it is incredibly common, are mixed and causal evidence is lacking.

Accreditation is prevalent in the civilian healthcare system despite the lack of compelling evidence on effectiveness because it is a prerequisite for receipt of reimbursement from Medicare (Jha 2018; The Joint Commission 2024). As the incarcerated population is stripped of federal benefits through the Medicaid Inmate Exclusion Policy, there is no such high-powered incentive for those operating U.S. jails. Nor is there a legal mandate for U.S. jails to obtain accreditation.

Another striking difference specific to delivery of health care in jails is the "paradox of having the same

[9] See *Miltier v. Beorn*, 896 F.2d 848 (4th Cir. 1990).

[10] In recent years, several states have received Section 1115 Waivers which allow Medicaid funds to be used for a limited set of health care delivered to the incarcerated prior to release (Saloner 2023; Howell et al. 2023).

[11] Data on the privatization of health and other services in U.S. jails are difficult to find, but estimates suggest that approximately 70% of jails contract with a for-profit health care vendor, such as Wellpath or Corizon, with these vendors often backed by private equity firms (McLeod 2019; Coll 2019).

[12] Compared to other quality assurance mechanisms such as warranties or licensing, quality disclosure systematically measures and disseminates information about product quality, is typically conducted by a third-party certifier, and relies on standardized quality standards so that results are comparable across producers (Dranove and Jin 2010).

institution meting out punishment and providing care for health problems" (Prout and Ross 1988). As experts noted when first developing standards in the wake of *Estelle*, the correctional setting means that "general schedules that strictly regulate work, exercise, and diet necessarily collide with individual medical orders for treatment" (Neisser 1977). Wishart and Dubler (1983), for example, state that "in delivering care to inmates, prison health care staff are responsible to their patients while simultaneously being constrained by the setting designed to separate, confine, and punish. They must accommodate the different and often conflicting norms that govern health care providers on the one hand and correctional staff on the other." Indeed, our own qualitative interviews reveal this problem, with incarcerated individuals noting that custody staff do not always implement the medical recommendations of medical staff because "the officers are more concerned about the disciplinary and correctional aspects than whether they are receiving the proper care." Thus, accreditation for U.S. jails is not handled by the TJC given these unique features; rather organizations like the American Correctional Association (ACA) and National Commission on Correctional Health Care (NCCHC) serving as leading accrediting bodies.[13]

A natural question is why jails would voluntarily undergo an accreditation process when it is neither mandated nor financially incentivized. In an original survey we conducted in 2020 (prior to the current study), we asked nearly 200 jails to provide reasons for why they either chose or were interested in undergoing accreditation. Reducing liability, obtaining feedback from experts, and improving health outcomes were all common rationales – with the most frequent reason (59%) being signaling a constitutionally acceptable level of care.[14] In particular, a decrease in grievances and lawsuits could result in a reduction of liability insurance premiums, with some accreditation agencies advertising that the receipt of accreditation can reduce premiums by up to 10%.[15] Responses were similar among the set of jails not yet accredited, although 32% of the non-accredited facilities indicated that they could see no reason to becoming accredited in the future. Other non-accredited jails noted that the cost of accreditation and lack of staff were binding constraints.

III Conceptual Framework

Appendix Section B formalizes the potential effects of accreditation. The setup is a principal (the sheriff) who seeks to minimize excess (greater than expected) inmate mortality. There are two agents – custody and medical staff who must work together to reduce the "aggregate output" of inmate deaths (Alchian and Demsetz 1972; Holmstrom 1982). Agents simultaneously choose one of two effort levels (high vs. low) and are one of two types (t^1 and t^2), with type t^1 agent using a state-dependent strategy and t^2 agent choosing low effort. Effort across agents exhibits perfect complementarity. Each agent knows his own type but is uncertain about the other agent's type and, in the baseline case, the sheriff is unable to observe individual effort. This static Bayesian game of incomplete information has a unique equilibrium prior to accreditation where both agents of both types choose low effort (*i.e.*, shirk) and death rates are excessively high.

[13] TJC briefly attempted to do correctional health care accreditation before abandoning the idea (Chassin, Mark R. 2021).

[14] Specifically, among 75 already accredited jails, 59% indicated that accreditation "signals a constitutionally acceptable level of care," 43% indicated "fewer grievances and lawsuits," and 47% stated that the accreditation process would "provide feedback from knowledgeable authorities." With respect to health outcomes, 55% of jails stated that accreditation would "improve health status of current inmates" and 44% said it would "reduce health risk to the community when inmates are released."

[15] See https://www.aca.org/ACA_Member/ACA/ACA_Member/Standards_and_Accreditation/Seeking_Accreditation_Home.aspx. Conversely, our own conversations with jail underwriters suggest that insurance premiums for off-site/catastrophic inmate medical insurance might increase in response to accreditation if medical claims rise because facilities may refer out inmates for medical care more often because of accreditation.

Given this setup, the accreditation process could lead to the following scenarios:

1. **Coordination**: Accreditation could enable agents to coordinate on the Pareto-efficient equilibrium, where both agents choose high effort. This can occur through two channels:
a. Horizontal Coordination: Accreditation could serve as a perfect signal informing the agent of the other agent's type. Since the game is symmetric, this introduces a second pure strategy equilibrium where both agents exert high effort.

b. Vertical Coordination: Accreditation could also serve as a contracting device, allowing the principal to perfectly observe each agents' actions. In this scenario, the sheriff could levy agent-specific penalties for exerting low effort, resulting in a unique Pareto-efficient high-effort equilibrium (Crémer and McLean 1988).

2. **Rubber Stamp**: If accreditation is merely a rubber stamp, no penalty is levied for whichever effort level an agent chooses. In such a scenario, both agents continue to put in low effort. In a richer model of social welfare, firm profits (for the accrediting agency) might rise and the risk of legal action (against the sheriff) may fall at the cost of taxpayer-financed subsidies for accreditation.

3. **Distraction**: Accreditation improves outcomes only if the high effort dictated by standards meaningfully improves health outcomes. In the distraction scenario, the standards are incorrect – leading to waste or harm. Each agent is perfectly informed about the other's type but are induced to choose low effort since high effort does not meaningfully lower inmate mortality.

This straightforward framework generates testable predictions. For example, the rubber stamp scenario implies that one might not observe any changes in compliance with health care standards. Under the distraction scenario, compliance with health care standards could increase, but mortality may exhibit no change. But if accreditation results in better horizontal and/or vertical coordination, one would anticipate improvements in both compliance with health care standards and mortality. We provide evidence on mechanisms via staff surveys below.

IV Intervention and Experimental Design

In this section, we provide a brief background of our intervention, experimental design, and implementation. Further details are available in Appendix Sections C, D, E, and F.

IV.1 Intervention

In this study, we randomize jails to undergo the NCCHC accreditation process. NCCHC grew out of the advocacy efforts of the American Medical Association (AMA) and American Bar Association (ABA) and is one of the oldest existing health care accrediting agencies in the U.S.[16] NCCHC standards are often referred to as the "gold standard" in correctional health care and leading policymakers have advocated for mandating accreditation, such as the NCCHC process, for all correctional facilities (Berwick, Beckman and Gondi 2021; Fiscella, Beletsky and Wakeman 2017). Courts also rely on NCCHC accreditation as a signal of adequate health care that complies with the Eighth Amendment. For example in *Balla, et al. v. Idaho State Board of Correction*, a federal judge noted that "The NCCHC accreditation..., while not determinative, constitute substantial evidence of adequate medical care."

[16] See Appendix Sections L.2 and L.3 for further history.

Today, NCCHC administers consensus-driven, proprietary standards covering seven major categories: (1) governance and administration; (2) health promotion, safety, and disease prevention; (3) personnel and training; (4) ancillary health care services; (5) patient care and treatment; (6) special needs and services; and (7) medical-legal issues. For example, governance and administration standards address the establishment of a health care system that ensures access to care, professional administration of all aspects of health care, and monitoring and quality improvement policies that effectively process health care issues from identification through resolution. As another example, patient care and treatment standards are designed to ensure the delivery of health care from arrival through discharge in a timely and appropriate manner. Specific examples of standards within each category can be found in Table 1. Each standard is also classified as either "essential" or "important" (see Appendix Table A1 for examples).

These NCCHC standards reflect the centrality of collaboration between correctional and medical staff. As NCCHC notes, "the primary mission of custody staff is to maintain order and secure the environment. The primary mission of health care staff is to identify, assess, and treat individuals' health needs" (National Commission on Correctional Health Care 2021*b*). As a result, "the relationship between custody and health staff can at times seem almost adversarial. All too often, it's 'us vs. them.' But health and custody staff are on the same team, and 'playing well together' benefits everyone – most importantly, the incarcerated patient population" (National Commission on Correctional Health Care 2021*a*). Collaboration is directly addressed in various NCCHC standards. For example, the NCCHC standard on medical autonomy states that "delivery of health care in a correctional facility is a joint effort of custody and health staff and is best achieved through trust and cooperation." Another NCCHC standard requires health training for correctional officers so that they can "recognize the need to refer an inmate to a qualified health care professional."

Figure 1 depicts the timeline of the NCCHC accreditation process, with further details provided in Appendix Section L.3. After a facility applies for accreditation, the NCCHC sends the facility the standards with a self-survey questionnaire that the facility can use to determine compliance and begin improvements. NCCHC then conducts an on-site visit, during which time NCCHC staff and physician surveyors review custody and medical records and speak with facility staff and patients. Findings from this on-site visit are incorporated into a report that is then presented to the Accreditation Committee. This Committee consists of physicians and nurses who are affiliated with professional societies (*e.g.*, AMA), industry (*e.g.*, Naphcare, health care executives) and government (*e.g.*, local county jails). The Committee is blinded to the facility's identity and determines compliance with standards based on the report.

To receive accreditation, a facility must be compliant with 100% of applicable essential standards and at least 85% of applicable important standards. If a facility is compliant with more than 90% but less than 100% of applicable essential standards and at least 85% of applicable important standards, the facility receives a status of "continuing accreditation with verification (CAV)." CAV status requires a facility to undergo corrective action. If corrective action is completed within the required four month time frame and verified by the Committee, the facility then receives full accreditation. Other facilities who are farther from the compliance thresholds are placed on "provisional accreditation" or "probation," which automatically requires another on-site visit (*i.e.*, focus survey). Facilities who require these visits are then re-reviewed by the Committee. Facilities that do not submit corrective action in a timely manner, or fail to resolve compliance findings, are placed on probation and those that remain non-compliant are ultimately denied accreditation. As we will discuss below, 41% of our treated facilities did not attain accreditation. Thus, we use an instrumental variables approach that either captures whether or not accreditation was completed or the fraction of the

accreditation process completed (measured either by NCCHC or the research team at Harvard).

IV.2 Study Timeline and Randomization

Figure 2 shows our study timeline, with the study commencing in early 2020 and ending in Fall 2024.

Recruitment and Enrollment: The study began with the recruitment of jails at national meetings and conferences targeting sheriffs and jail administrators (*e.g.*, Committee of State Sheriffs, the National Sheriffs' Association (NSA), the American Jail Association (AJA), and NCCHC). A member of our research team would attend either in person or virtually. For in-person meetings, team members set up vendor booths with informational flyers and business cards for our project coordinators.

Sixty interested facilities expressed initial interest in the study and we conducted an informational session with custody and/or medical leadership from each interested jail. During the informational session, we shared additional details about the study and information on our Certificate of Confidentiality from the National Institutes of Health (NIH) which we obtained to further protect the information provided by participants. See Appendix Figure A1 and Appendix Section C for material and further information.

Jails were eligible for our study if they had an average daily population (ADP) of between 100 and 3,000 (midsized jails) and had not previously received accreditation by the NCCHC. Privately operated jails, ICE facilities, and juvenile detention facilities were excluded from the study. Out of the 60 interested facilities, 12 declined to move forward following the informational session, one did not meet our study eligibility criteria, and one was not eligible for accreditation based on existing NCCHC standards.

To formally enroll in the study, we required written and informed consent signed by a representative of both custody and medical staff. Following enrollment, eligible jails were provided our baseline facility and staff survey. Completion of the facility survey was incentivized by $500, which could be used towards the cost of accreditation. Staff surveys were incentivized by a lottery of winning a $100 e-gift card.

All enrolled facilities were given the opportunity to become accredited at a highly subsidized rate, with the control facilities beginning the NCCHC accreditation process at the end of the study. Accreditation fees are based on jail ADP. The subsidy for the treated group was much larger and available immediately compared to the control group, with a difference of about $5000 (Appendix Table A2).

Randomization: Once jails completed both baseline surveys, they were randomized into the treatment ("Accredit Now") or control ("Accredit Later") arm and informed of their assignment. Stratified randomization was conducted based on ADP and timing of enrollment ("batch"). Given that larger facilities might have more resources and/or more health challenges, we stratified on high versus low ADP to improve statistical power (Athey and Imbens 2017; Bruhn and McKenzie 2009). Because there were capacity constraints in the accreditation process, we also randomized in three batches, thus yielding six total strata.

NCCHC Accreditation Starts: After randomization, NCCHC begin its accreditation process for the treatment facilities, with treatment compliers (*i.e.*, treated facilities that took up treatment) undergoing the NCCHC accreditation process depicted in Figure 1. NCCHC did not deviate from its usual process for the purposes of our study. Once the treatment facilities received an initial accreditation decision from NCCHC, both control and treatment groups completed incentivized endline facility and staff surveys, as well as medical audits and inmate and leadership interviews. Note that some facilities had still not received NCCHC accreditation by the time of our Harvard endline since they were still going through corrective action. These survey

instruments are described in Section V.1 below.

IV.3 Sample

Our initial goal was to recruit 40 facilities with a wait list for additional interested facilities. We invoked randomization off the wait list after two facilities randomized to treatment were removed for reasons orthogonal to their assignment status: in one instance the consenting party suffered a heart attack and the other was put on leave for sexual assault. We included six additional facilities from the wait list in our sample, randomizing such that 22 facilities were allocated to treatment and 22 to control in our main analytical.[17] See Figure 3 for a diagram of the number of interested, enrolled, randomized, and treated facilities in the study. Appendix Table A4 shows that the 44 enrolled facilities are similar to the facilities that declined to participate in terms of baseline ADP, health vendor status, sheriff reelection status, and mortality.

IV.4 Representativeness

Our recruitment efforts yielded a sample of jails from all over the U.S., including regions where incarceration rates are highest (Zeng and Minton 2021). See Figure 4, which illustrates where the 44 facilities in our main analytical sample are located, with each circle representing the total ADP across all facilities in each state. We use data from the 2019 Bureau of Justice's Census of Jails (U.S. Department of Justice 2022) to assess the representativeness of our study sample. The Census is designed to capture all jails although reporting is not mandatory.

Comparing the 41 out of 44 jails in our sample who responded to the 2019 Census of Jails against 1,085 other medium-sized (ADP 100-3,000) county jails serving adult populations, we find that our sample is representative of these other jails in terms of jail characteristics (Table 2 Panel (A)) and characteristics of the incarcerated population (Panel (B)), although jails in our study had slightly higher yearly admissions and a lower share of White incarcerated individuals. Approximately 4% of our sample and similar out-of-sample U.S. jails were under a consent decree for conditions of confinement and about half of all in-sample and out-of-sample jails held felony offenders. Across all jail and incarcerated population characteristics, the overall F-stat is 1.334 with an associated p-value of 0.2.

While broadly representative of medium-sized jail in the U.S., the jails in our main sample are likely negatively selected relative to other jails that voluntarily seek NCCHC accreditation. Based on information provided by NCCHC, the vast majority of jails who apply for accreditation eventually receive accreditation although a small set of extremely negatively selected facilities may also be court-ordered to obtain accreditation as part of a consent decree or monitoring process (Douds and Ahlin 2016).

Our recruitment efforts revealed many jails had never heard of NCCHC. And among our sample of 22 treated facilities, only 13 successfully received accreditation. Anecdotally, the Accreditation Committee noted to the research team that they were surprised by the lower compliance and success rates of ''Harvard facilities" compared to the typical facilities that apply for accreditation. Although the "Harvard facilities" had little information about NCCHC prior to their recruitment, each of the 13 treatment facilities that received accreditation announced their status on jail websites or social media accounts.

[17] Appendix Table A3 shows that treatment assignment is not a statistically significant predictor of involuntary removal. Neither facility initiated accreditation, and one never connected with the NCCHC. We impute endline outcomes for the two non-initiating facilities using their baseline values and find similar results to those presented in Section VI.

V Data

This section describes our survey instruments, coding methodology, and construction of indices for analysis. We also discuss steps taken to ensure the accuracy and reliability of our measured outcomes.

V.1 Survey Instruments

We designed four primary survey instruments for the purposes of this study: (1) facility survey, (2) anonymous staff survey, (3) medical audit, and (4) qualitative interviews with incarcerated individuals and facility leadership. We also conducted an expert survey to obtain (a) weights on the importance of each quality standard and (b) baseline features of facilities predicted to impact the effect of accreditation. Details on the data, along with links to all survey instruments, are provided in Appendix G.

In designing the quantitative survey instruments, the research team analyzed the NCCHC quality standards to determine (a) whether and to what extent each standard could be objectively assessed and (b) to determine the best manner to obtain accurate information. For example, NCCHC's Privacy of Care standard requires that health care encounters and exchanges of information are conducted in private (*e.g.* private area, privacy screen, curtain). To measure whether a facility complies with this standard, our facility survey required the facility to upload photo documentation. As another example, NCCHC's Receiving Screening standard requires medical personnel to inquire about illness histories, allergies, recent communicable symptoms, prescription medications, drug use, and other conditions shortly upon intake. In our medical audit, our physician surveyor visually inspected randomly selected medical documentation of receiving screenings in order to assess compliance. Our anonymous staff survey, which is not part of NCCHC's accreditation process, also enabled us to collect systematic information on staff sentiment and cross-validate the other metrics.

(*1*) *Facility Survey*: The facility survey was administered at both baseline and endline via a secure Qualtrics link for each facility. The facility survey had two major components: a questionnaire that was answered by the facility administrator, and document uploads of requested procedures, protocols, and minutes from specified meetings. In the questionnaire component which included 67 question items, facilities were asked about a range of topics spanning the quality standards, such as the availability of specific physical equipment and medical supplies, the availability of mental health and dental services, the FTE of each type of medical staff, and how incarcerated individuals were screened upon booking and for specific conditions. The upload component of the survey required facilities to provide a copy of their official polices and procedures manual, along with 43 other requested documents. These included documentation on management of COVID-19 and protocols for Medication Assisted Treatment (MAT), treatment protocols for patients with conditions such as hypertension, diabetes, and HIV, as well as CPR and other health certification for all personnel. In addition, we requested blank copies of health assessment forms, mental health screenings, sick call forms, and death logs, among others. Finally, we requested health staff meeting and administrative meeting minutes and attendance records from the past 12 months.

(*2*) *Staff Survey*: The staff survey was also administered at both baseline and endline via a secure Qualtrics link for each respondent. The survey was sent to comprehensive staff email lists that included both custody and medical staff. We asked staff approximately 120 questions related to staff satisfaction, the relationship between custody and medical staff, and perceptions of care.

(*3*) *Medical Audit*: The medical audit was administered at endline with a retrospective look back to the study

baseline period. The audits were performed on Zoom with a medical professional alongside a scribe and took approximately three hours per facility. For the audit, each facility was asked to pull a random set of medical records for patients with the following conditions: suicidality, opioid use disorder, and hypertension, as well as additional randomly selected files, covering both endline and baseline periods. In addition, facilities provided death logs (which include date and cause of death for individuals who died while incarcerated), clinical death reviews, and psychological autopsy reviews for the past 12 months which were reviewed and documented by the research team. All records (whether electronic or paper) were shown to the research team virtually, with personally identifiable information redacted.

(*4*) *Interviews with Incarcerated and Leadership:* We also conducted qualitative interviews with incarcerated individuals and facility leadership at endline only. For each facility, we requested interviews with three incarcerated individuals (at least two of whom had used medical services) and with at least one senior facility administrator. These interviews were conducted on Zoom, took approximately 45 minutes each, and were conducted in private with the study coordinator and a scribe. Incarcerated individuals were asked questions in both multiple choice and open response format regarding their perceptions of the care they received. Administrators were also asked questions about their perceptions of health care delivery. Unlike the other surveys, these interviews were not considered mandatory for the facility to complete and the coverage rate is accordingly lower.

(*5*) *Expert Survey:* From October 2022 to March 2024, we collected incentivized predictions on our study from 145 experts across various domains, including jail administrators and health care providers, policy/academic experts, and the formerly incarcerated. Recruitment occurred at national subject matter conferences, through cold calls of expert lists, and using recruitment methods on Prolific. We use expert weights to construct an aggregate quality standards index, as described below. Respondents were asked to rank order from 1–7 a set of seven outcomes in terms of perceived importance. We also take a principled approach to assessing heterogeneity by relying on predictions from this survey: Respondents were asked to select five factors that they believed would most affect whether accreditation of jail health care services by NCCHC has an effect on outcomes. We use these predictions in our heterogeneity analysis as described below. See Appendix G.6 for further details on the expert survey.

V.2 Additional Data

We use additional data from the Census of Jails and also scraped data on mortality events in the media. In addition, we trained law students to code information on lawsuits, such as incident dates and whether the legal claim alleged inadequate health care, using case filings and decisions from LexisNexis, CourtListener and PacerMonitor. Each facility was coded by two independent law students who were blinded to the treatment assignment. If there was disagreement between two coders, a third independent law student would adjudicate. Finally, we obtained jail bookings records and scraped publicly-available jail rosters to measure recidivism (defined as rebooking in the same jail) three and six months post-release.

V.3 Coding and Construction of Main Outcomes

Our main outcomes are (1) staff sentiment, (2) quality standards compliance, and (3) mortality. Details on the coding and construction of these outcomes can be found in Appendix H and Appendix I.

(*1*) *Staff Sentiment:* We measure staff sentiment using individual-level responses in the staff survey. All ques-

tions utilized five point Likert scales, which we rescaled to fall in the unit interval. We then created several indices to measure staff sentiment. We obtained endline responses from 40 jails and we impute missing endline responses for the other 4 facilities in our analytical sample with baseline values. For example, our horizontal coordination index represents the arithmetic average of questions about custody supporting clinical decisions by medical staff, collaborative work between the two, and open communication. Our vertical communication index consists of questions on supervisors being open to concerns about the health or safety of inmates and staff and whether staff feel comfortable bringing up concerns to their supervisor. We also create an index of job satisfaction and assess staff views on respect for incarcerated individuals.

(2) *Quality Standards:* To measure compliance with quality standards, we use data collected from the facility survey and medical audit, which were mapped into each of the seven NCCHC quality standard categories. We used a specific coding process to independently and objectively measure compliance on the basis of the documents uploaded by facilities in the facility survey. Specifically, a Coding Committee consisting of Harvard students and chaired by the study coordinator developed a coding manual to map raw data elements into a dataset. The manual consists of 43 separate spreadsheet pages corresponding to the major compliance indicators within each of the seven quality standards. Each sheet provides a step-by-step guide for coders to determine whether the facility is complying with the standard. As one example, the current NCCHC standards require facilities to have a grievance policy for incarcerated individuals that provides clear guidance on days to first response, formal documentation of grievances, and an appeals process. The coding manual instructed our coders to read uploaded protocols to determine whether such criteria was met, as can be shown in Figure 5.

For each facility, this coding process was done by two independent college or law students who were blinded to the treatment assignment. Students were only eligible for coding after undergoing an online training session and practice exam which included intentional errors to ensure their accuracy and attention to detail. If there was disagreement between the two coders as to whether a facility complied with a standard, a third independent student (again blinded to treatment assignment) would adjudicate. Another Coding Committee consisting of Harvard medical students, study coordinator and one of the PIs, mapped questions in the medical audit to each of the seven quality standard categories.[18]

At the end of the coding process, we created an index for each of the seven NCCHC standards categories. All elements in each index ranged from 0 to 1, with either binary indicators to measure compliance with the standard, or a fraction to indicate percent compliance when a standard required multiple elements. Non-response was penalized as described in Appendix Section H.2.2. We take the index for each quality standard as the arithmetic average across all such indicators associated with that standard. Thus, each index can be interpreted as the aggregate percent compliance with the standard, with 0 indicating complete non-compliance and 1 indicating full compliance.

(3) *Mortality:* We use death logs obtained from the medical audit to measure mortality, with the time range of the death logs covering the prior 12 months. To independently verify that facilities were not under-reporting mortality, coders collected deaths reported in the media, which were then matched to death logs using the date and cause of death. We then took the union of deaths from both the medical audit and the media, dropping duplicates, to ensure that we were capturing all identified deaths. For four facilities who did not provide data logs (all in the control group), we only use deaths reported in the media. Using the union of deaths from the medical audit and media reports, we calculate the average deaths per month using a

[18]Further description on all coding processes is provided in Appendix H.

baseline period that covers six months prior to randomization and an endline period covering six months after the expected time of NCCHC's on-site visit.[19] The correlation between endline mortality based on death logs and media reports is 0.82, and we explore non-linear specifications and measures in robustness checks.

V.4 Aggregation of Quality Standards

As per our pre-analysis plan, we aggregate the seven indices representing each category of quality standards in several ways to obtain a meta-index. First, we give all seven indices the same weight and calculate the simple average across all indices. Second, we use a principal component analysis with all outcomes (214 in total), using the first component to create a meta-index. One caveat is that the weights with which the outcomes are aggregated to components are not necessarily positive. We rescale the first principal component so it lies in the [0,1] interval (Lise and Postel-Vinay 2020).[20] Finally, we calculate a weighted average across the seven indices where weights are determined by asking experts (*e.g.*, health care, corrections, policy experts, and previously incarcerated individuals, detailed above in Section V.1) the importance of different outcome categories. The reciprocal rank weights are presented in Figure 6 and show that experts typically rank standards on patient care and treatment as most important, with less weight on governance and medical-legal standards.[21]

V.5 Measurement

Our survey instruments rely on documentation and information provided by facilities who were informed of their treatment assignment, raising the possibility of demand bias or Hawthorne effects. We take several steps to ensure the integrity of the data we collected.

As described above, we designed our survey instruments to target quality standards that could be objectively measured and verified by the research team. For example, we measure compliance with quality standards using components from the facility survey using both facility "self-reported" answers but also independent coding of uploaded documents by students, as well as data from the medical audit. Figure 7 Panel (A) demonstrates the composition of our seven indices by data source, illustrating substantial variation in the extent to which a given standard relies on each source. Notably, quality standards on "safety and prevention" and "patient care and treatment" utilize a high share of data from the medical audit, which relied on direct observation and inspection by our medical personnel and scribe, and rely minimally on facility self-reports. Thus, we view the measurement of these standards as objective and difficult to falsify. We also intentionally chose to exclude staff survey responses in our measurement of compliance with quality standards. Instead, we use the staff survey to create separate metrics of quality and outcomes.

[19] The mean time from randomization to on-site visit was 9.5 months, so our post-treatment period covers the 10-15 months post randomization.

[20] Anderson (2008) proposes an algorithm to aggregate outcomes to an index using the inverse covariance matrix for weighting. Our preliminary power analysis found that indices created with this algorithm perform poorly when a large number of outcomes are to be aggregated.

[21] Specifically, experts ranked standards from most important ($r_i = 1$) to least important ($r_i = 7$) and these ranks are transformed to weights w_i as follows: $w_i = \frac{\frac{1}{r_i}}{\sum_{j=1}^{n} \frac{1}{r_j}}$. The formerly incarcerated (lived experience experts) also place more weight on personnel training whereas the academics, administrators and health care providers (professional experts) place more weight on safety and prevention. Our main aggregation index combines both sources of expertise. Further details on our aggregation methods are provided in Appendix I.

Second, we do not use any of the data NCCHC collected in its on-site visit or final report to the Accreditation Committee as our outcomes. All of the data used in this study to determine whether facilities are meeting standards were collected and coded independently by the research team at Harvard. Figure 7 Panel (B) compares our study's compliance score for all essential standards with NCCHC's compliance score.[22] Recall that a facility must meet 100% compliance with essential standards in order to receive NCCHC accreditation. While the two measures are highly correlated at 0.71 (p-value = 0.0004), Harvard compliance scores are substantially lower on average, with average scores of 0.58 and 0.74 for the study and NCCHC, respectively.

Lastly, we verify the accuracy of our mortality measure given potential concerns about underreporting of deaths. As described above, we cross-check each death reported in the death logs with media reports. In general, we find that the death logs are more complete than media reports because not all deaths are released to news outlets. Our main mortality outcome includes the union of all deaths from either log or media sources, but our findings are robust to using only deaths reported in the news.

V.6 Summary Statistics and Balance

Table 3 presents baseline means of key variables and assesses for balance between treatment and control facilities in our main analytical sample. Panel (A) reports baseline compliance with quality standards, Panel (B) reports baseline mortality, Panel (C) reports baseline staff survey responses, and Panel (D) reports baseline jail characteristics. The first column reports the overall sample mean, the second column represents the coefficient on the treatment assignment indicator, and the third column reports the associated standard error, adjusting for randomization strata fixed effects.

Panel (A) of Table 3 shows that at baseline, jails in our study meet 45-54% of quality standards on average, with the lowest rate of compliance for "patient care and treatment" and the highest rate of compliance for "medical-legal issues." Importantly, there are no significant differences between treatment and control facilities on any of the standards. Panel (B) shows similar balance in the baseline mortality rate.[23] Panel (C) shows that baseline staff survey responses are also similar both in terms of number of responses, respondent demographic characteristics, and baseline measures of staff sentiment indices such as horizontal coordination (between the medical and custody staff) and vertical coordination (between staff and supervisors). Panel (D) demonstrates similar balance on baseline jail characteristics such as number of staff (FTE), whether the jail utilizes a for-profit health vendor (68%), whether the jail was previously accredited by a non-NCCHC organization, and political affiliation of the county and sheriff. The overall F-test of all baseline characteristics is 1.656 with an associated p-value of 0.203.

Figure 8 presents additional descriptive evidence on the deaths among facilities in our main analytical sample. Panel (A) presents the distribution of deaths by cause of death. The most common cause of death in jails is undetermined (35%) followed by suicide/homicide (28%), illness (21%), and drug overdose (13%). While some of these undetermined deaths may still be actively under investigation, our independent review of media reports associated with these deaths indicates that many undetermined deaths involved situations alleging homicide, medical emergencies, and drug overdose. The prominence of deaths due to natural

[22] This comparison is only feasible for treatment facilities because NCCHC does not compute scores for control facilities as they do not interact with them until study completion.

[23] We note that this baseline uses media-reported deaths only as we did not obtain death logs that extended far enough back in time to use death logs consistently across facilities. However, the experimental findings on mortality are robust to using death reported in the media only.

causes (21%) is concerning given the average age of incarcerated individuals in the sample is 33. Panel (B) of Figure 8 demonstrates a marked seasonality of deaths in jails – with a spike in the third quarter when 40% of all deaths occur. These patterns are consistent with lack of air conditioning and other heat-mitigation systems that may worsen chronic conditions and/or contribute to acute dehydration and heat stroke (Skarha et al. 2022, 2023; Cloud et al. 2023; Tuholske et al. 2024; Cowan, Lemasters and Brinkley Rubinstein 2024) and increase violence while incarcerated (Mukherjee and Sanders 2021).

VI Empirical Strategy and Results

In this section, we describe our empirical strategy before turning to our main results on compliance with staff views, quality standards, and mortality.

VI.1 Empirical Strategy

We estimate the reduced form intent-to-treat (ITT) effect of assignment to treatment (initiation of the NCCHC accreditation process) using an OLS regression specification of the form:

$$Y_j = \alpha_0 + \alpha_1 Treat_j + X_j\Gamma + \mu_j + \epsilon_j \tag{1}$$

where j indexes facility. Y_j comprise our key outcomes of compliance with quality standards, mortality, and staff sentiment. $Treat_j$ is an indicator for being randomly assigned to the treatment group and μ_j represent randomization strata fixed effects. Baseline measures of the relevant outcome, X_j, are included in some specifications. We use heteroskedasticity-robust standard errors throughout for facility-level outcomes and cluster standard errors at the facility level for individual-level staff outcomes. We also compute *p*-values using randomization inference (RI) obtained by permuting accreditation assignments for each facility 1,000 times.

As noted, some treated facilities failed to complete the accreditation process. We use two-stage least-squares to obtain the treatment on the treated (TOT) estimate:

$$\begin{aligned} Y_j &= \alpha_0 + \alpha_1 \widehat{Accred}_j + X_j\Gamma + \mu_j + \epsilon_j \\ Accred_j &= \beta_0 + \beta_1 Treat_j + X_j\Gamma + \mu_j + \epsilon_j. \end{aligned} \tag{2}$$

where $Accred_j$ is an indicator for whether accreditation was successfully completed according to Figure 1 and treatment assignment is used as the instrument. Control complier means are reported throughout (Kling, Liebman and Katz 2007).[24] The 2SLS estimates account for imperfect compliance as only 13 out of 22 facilities will have received accreditation, with most failing facilities either not implementing any of NCCHC's recommendations and/or not agreeing to a required focus survey. In robustness tests, we code as the endogenous variable the fraction of the accreditation process completed, to allow for the possibility that the scaling is inaccurate, (*e.g.*, if treatment non-compliers still benefited from undergoing some portion of the process). We determine the fraction of completion rates using either NCCHC metrics or the Harvard study metrics.

[24]Specifically, we calculate control complier means using two-stage-least-squares regressions of $Y_j \cdot \mathbf{1}[Accred_j = 0]$ on $\mathbf{1}[Accred_j = 0]$ instrumenting with $Treat_j$ along with baseline controls and randomization strata fixed effects.

VI.2 Staff Sentiment

We begin by assessing the effects of our intervention on staff sentiment given the predictions of our model in Section III. Reduced form estimates from Equation 1 are presented in Figure 9. The gray bar represents the regression-adjusted control mean and the blue bar represents the control mean plus the estimated treatment effect. These reduced form estimates show large and statistically significant improvements on indices of horizontal coordination, with more modest gains in job satisfaction.

Corresponding two-stage-least-squares estimates from Equation 2 are shown in Table 4. Second stage estimates are presented in Panel (A) and Panel (B) shows first stage *F*-statistics, which range from 56 to 106 depending on the specification. Our 2SLS estimates indicate that completing the accreditation process increases horizontal coordination by 5.9 percentage points, a 10% increase from the control complier mean. Staff satisfaction increased by 3.7 percentage points, a 6% increase. In contrast, vertical coordination and staff views on respect for incarcerated individuals were unaffected by accreditation, though aspects of vertical coordination such as whether staff feel comfortable speaking up on inmate care issues are positively affected (Appendix Table A5).

Appendix Table A5 reports results for each individual component of these staff sentiment indices. These results show that completing the accreditation process increased many components such as whether custody and medical staff coordinate and communicate, whether staff feel they are properly trained and whether they feel valued at work. These findings support the horizontal coordination case in our model, which should portend improvement in both compliance with quality standards and mortality, if the standards are aimed appropriately (*e.g.,* not a distraction). These results also echo what we find in the qualitative comments: where many in facility leadership noted the importance of knowing what the correct processes were and being able to obtain buy-in from custody staff.

Did accreditation lead to an increase in the number of staff, use of staff (*i.e.,* change in referral patterns), health care management or capital investments? Table 5 suggests it did not. The main differences between treatment and control facilities were in how the staff cooperated with each other and the services (*e.g.,* lab and pharmacy) they provided to the incarcerated individuals. We next turn to whether these changes resulted in differences in the quality of care, as measured not by staff surveys, but by facility surveys and medical audits.

VI.3 Quality Standards

Meta-Index of Quality Standards: We assess the impact of our intervention on a meta-index of quality standards, as described previously in Section V.4 and per our pre-analysis plan. Reduced form estimates from Equation 1 are shown in Figure 10, which present compliance with the meta-index constructed in three ways: (1) a simple average across the seven quality standards, (2) a principal component analysis (PCA) using all outcomes as inputs and taking the first principal component as the meta-index, rescaled to the unit interval, and (3) a weighted average of the seven quality standards using the reciprocal rank weights from our expert survey. These reduced form estimates show large and statistically significant improvements on the meta-index using PCA and expert weights.

Table 6 shows corresponding two-stage-least-squares estimates from Equation 2 on these meta-indices constructed using the three approaches. While the simple average weighting scheme yields a positive but insignificant effect of completing the accreditation process, both PCA and expert schemes – which place

more weight on the more objective standards including patient care and treatment – yield significant improvements in the meta-index of quality standards. For example, expert weights yield a 8.5 percentage point increase in quality standards, a 15% increase relative to the control complier mean.

Appendix Table A6 presents results using different expert weights for formerly incarcerated respondents (who are the recipients of care) and all other experts (who are typically providers of care) and shows that our findings are robust to either weighting scheme.

Seven Quality Standard Indices: Figure 11 presents reduced form estimates from Equation 1 on the more disaggregated seven quality standards. As before, the gray bar represents the regression-adjusted control mean and the blue bar represents the control mean plus the estimated treatment effect. Across all seven categories, facilities assigned to treatment improved their compliance relative to the control group although some changes were statistically insignificant. Compliance with "personnel training" and "patient care and treatment" improved by a statistically significant 15% and 11%, respectively, relative to the control mean.

Table 7 presents corresponding two-stage-least-squares estimates from Equation 2. Second stage estimates are presented in Panel (A) and generally mirror the reduced form patterns previously discussed in Figure 11. Specifically, completing the accreditation process increases compliance with "safety and prevention" standards by 9.8 percentage points, a 17% increase relative to the control complier mean. Completion of the process also increases compliance with "personnel training" by 14.3 percentage points, a 27% increase relative to the control complier mean, and increases "patient care and treatment" compliance by 10.2 percentage points, an 18% increase. Panel (B) shows first stage F-statistics, which range from 25 to 47 depending on the specification.

Appendix Figure A2 presents two-stage-least-squares estimates on substandards within each of the three categories we find improvements in. We observe increased compliance along the majority of substandards, with notable increases in compliance with clinical prevention (which includes testing for sexually transmitted infections), contraception, personnel credentials, receiving screening, mental health screening, and nursing procedures.

Screening: Based on our audit of inmate medical records, we also observe notable increases in documentation for health conditions and general and specific screening for medical conditions. Appendix Figure A3 presents two-stage-least-squares estimates on required components for screenings. For the receiving screening, which should be conducted within 24 hours of booking, we see increased screening of characteristics such as dietary needs, appearance, behavior, state of consciousness, breathing, and withdrawal symptoms (Panel (A)). We also see increased screening of mental health conditions within 14 days of admission, with additional screening for history of detoxification and outpatient treatment, and the status of psychotropic medications, suicidal ideation, orientation to person, place, and time, and emotional response to incarceration (Panel (B)).

Appendix Table A8 reports two-stage-least-squares estimates on the screening and treatment of inmates with suicidal ideation, opioid use disorder, and hypertension based on medical records for patients with these specific conditions. We observe increased screening for symptoms of suicidal ideation, including history of suicidal behavior and current status of suicidal ideation. Similarly, we also observe increased screening for symptoms of opioid use disorder, such as substance use hospitalization, withdrawal seizures, detoxification and outpatient treatment, etc. However, accreditation did not lead to a significant change in the treatment for any of these conditions. We find no evidence of increased monitoring or intervention for

inmates with suicidal ideation, education on medication-assisted treatment (MAT) or treatment for those with opioid use disorder, or specialized dietary or blood pressure management for inmates with hypertension.

VI.4 Mortality

Figure 12 shows the distribution of the endline mortality rate for the treatment and control facilities. The figure reveals a left shift in the mortality distribution for the treatment group relative to the control group, with treatment facilities concentrated at zero. A Kolmogorov-Smirnov test confirms that the distributions of the treatment and control groups are significantly different, with a p-value of less than 0.01. We also obtain a RI p-value of less than 0.01 when comparing treatment and control means controlling for randomization strata fixed effects.

Table 8 reports two-stage-least-squares estimates from Equation 2 on average deaths per month, again measured over six months following the expected time of NCCHC's on-site visit. Panel (A) presents second stage estimates and Panel (B) presents first-stage estimates, with F-statistics ranging from 27 to 34. Column (1) of Table 8 represents our baseline specification with randomization strata fixed effects only. Column (2) adds a control for the log ADP to proxy for jail population and column (3) is our preferred specification with both randomization strata fixed effects and baseline mortality. We find that completing the accreditation process significantly reduces mortality. In our preferred specification, for example, completing accreditation reduces average deaths per month by 0.25, representing a 86% decline relative to the control complier mean.

In Appendix Table A9, we report two-stage-least-squares estimates of the impact of completing accreditation on different parts of the mortality distribution. Accreditation increases the probability a facility had 0-1 deaths by 48.1 percentage points, but reduces the probability that a facility experienced 2-4 or 5 or more deaths by 32.3 and 10.0 percentage points, respectively. Given that mortality is right skewed and discrete, we also estimate nonlinear GLM specifications (Hardin, Schmiediche and Carroll 2003) using count models that follow a Poisson (columns (1)-(2)) or negative-binomial distribution (columns (3)-(4)) using $\log(\text{ADP})$ as an offset to capture the relationship between deaths and facility population. We also estimate a model of log mortality (columns (5)-(6)) and mortality per average daily population (column (7)). 2SLS estimates from these alternative specifications are similar to our baseline estimates (see Appendix Table A10). Results are also robust to using only deaths reported in the media (see Appendix Table A11). Appendix Table A12 presents two-stage least squares estimates of accreditation completion on total endline deaths by cause of death and shows reductions across all causes, with statistically significant reductions in deaths attributable to illness.

VI.5 Additional Outcomes: Lawsuits and Recidivism

Table 9 reports 2SLS estimates of accreditation completion on the outcomes of recidivism and lawsuits, with corresponding reduced form estimates presented in Appendix Table A13. We measure individual-level recidivism among 32 jails in our sample for whom we are able to obtain jail bookings data. Column 1 demonstrates that whether or not we could access such data is not correlated with treatment. Columns 2 and 3 reports results for three- and six-month recidivism among individuals released within a 45-day window following the endline survey. Accreditation is negatively associated with recidivism: Individuals released from accredited facilities are 12.8 percentage points and 20.8 percentage points less likely to be rebooked

in the same jail three months and six-months post-release, representing 54% and 52% reductions relative to control facilities, respectively. Columns 4–6 show effects on healthcare related litigation for 44 facilities. Estimates are generally negative but not statistically significant.

VI.6 Robustness

For all our main results, we assess robustness to the use of the starting sample – including the two facilities that never initiated the accreditation process – and to continuous versions of the endogenous variable – to address the possibility that the scaling in our two-stage least squares estimates is incorrect if treatment non-compliers benefited from undergoing some portion of the process.

First, Appendix Table A14 shows pre-treatment balance among the starting sample including those who never initiated accreditation. Appendix Tables A15, A16, A17, and A18 and Appendix Figure A4 present results for staff sentiment, quality standards, and mortality respectively among this sample, which are very similar to our main findings.

Second, we also estimate two versions of continuous improvement, one using the NCCHC scores for important standards, as defined in Appendix Table A1, and another using the Harvard measures of the same.[25] The first stage of the different estimation strategies is shown in Appendix Figure A5, which shows very little difference in the scaling across the varying approaches. The ITT implicitly scales by 1 and the binary TOT is a bit noisier compared to the more precise TOT continuous versions; however the magnitudes are indistinguishable suggesting scaling has trivial policy implications in our context. As shown in Figure 13, the findings for staff sentiment, quality standards, and mortality are similar to our main findings.[26]

VII Expert Predictions

We next turn to assessing heterogeneous effects using the factors predicted by experts. Recall that we conducted an expert survey of domain experts where respondents selected five factors that they believed would most affect whether accreditation of jail health care services by NCCHC has an effect on health care procedures, health outcomes, staff outcomes, and litigation. Appendix Figure A7 shows the ranking of these factors across all experts, with the top three factors being medical staff FTE per ADP, the quality of health care, and the quality of jail management.

Yet, we find little to no evidence that any of these factors predict heterogeneity. In Figure 14, we plot reduced form coefficients on the interaction between each factor that can be measured at baseline and treatment assignment on the endline expert-weighted meta-index of quality standards (Panel (A)) and endline mortality (Panel (B)). Although noisy, there is minimal heterogeneity in treatment effects across any of the factors. If anything, there is suggestive evidence that facilities located in a Republican county (the factor least likely to be predicted by experts to affect the impact of accreditation) experienced larger improvements in compliance with quality standards.

[25] Our Harvard endline measurement came before four treatment facilities finished their focus surveys. As these facilities did eventually receive accreditation, we set their Harvard scores to 90%, which corresponds to the important standards Harvard score attained by the treatment facility that became accredited without any further steps.

[26] This is also the case in the entire starting sample – see Appendix Figure A6.

VIII Qualitative Evidence

To better understand the nature of changes that occurred at treatment facilities as they underwent the NCCHC accreditation process, we now turn to qualitative evidence collected during our interviews with facility leadership. Among all treatment facilities, we asked several questions at endline on their experience with the accreditation process, challenges they faced, biggest change implemented, and the relationship between custody and medical staff.[27] Word clouds depicting responses to these questions are shown in Appendix Figure A8. This qualitative evidence supports our empirical findings which show that receiving accreditation improves staff coordination and compliance with quality standards (particularly patient care and treatment).

Experience Undergoing Accreditation: The vast majority of treatment facilities indicated that the process was "nervewracking," "stressful," "daunting," and "eye-opening." Many facilities indicated that "At the beginning, had no clue what were doing" and that this was the "first ever process for the history of this jail." Another common theme was that facilities needed to put in a lot of time and effort and described the NCCHC process as being "pretty thorough and intense." One facility noted that "NCCHC reviewers caught everything. Her ability to point out areas needing improvement was great." Another facility noted that it "was harder to convince corrections that this is something we should do than it was to go through the process. Accreditation process was helpful for highlighting changes that could be made to improve." Another facility commented that the "process in general is awesome... Its about finding shortfalls and closing the gap. Staff has become more involved in the medical process and have better understanding of why we are doing those things. Building relationships between medical and security was another piece."

Challenges Faced During Accreditation: Treatment facilities indicated a range of challenges. Several facilities mentioned difficulties having to do with changes in contracted health providers or staff turnover that occurred during the NCCHC accreditation process. Other facilities noted that it took a lot of work to change policies and obtain buy-in from staff: "Hardest part was just getting people on board." Obtaining support from custody staff, in particular, was a common theme. As one facility told us, "The biggest challenge was getting the officer buy in. Why are we doing this we have never done this before?" Several other facilities mentioned staffing and resources constraints with respect to meeting standards on suicide watch, chronic disease, and dental services.

Biggest Change Implemented: Almost all treatment facilities indicated that the biggest change was with respect to the aspects of screening and initial health assessments. One facility noted "It was that intake for the initial and it going immediately going to medical." Another facility stated that the "receiving screening is one of the bigger ones. Way more contact with patients when they come in now. Also coordination with custody staff is more comprehensive now."

Mental health screening and processes of care were also subject to substantial change, according to some facilities. For example, one facility stated, "Staying on top of mental health assessments was important. With the new protocols, the facility can ensure psychiatrists are staying on top of what they need to be doing." Other facilities shared similar changes, such as "Really focused on turnaround time for health assessments

[27] Specifically we asked facilities to (1) "Please describe your facility's experience undergoing NCCHC accreditation process," (2) "Were there any challenges your facility faced during the NCCHC accreditation process? If so, what were they?" (3) "In your opinion, what is the biggest change your facility has implemented as a result of the NCCHC accreditation process?" and (4) "Do you think that the relationship and coordination between correctional and health care staff has changed due to the NCCHC accreditation process? If so, how has it changed?"

and review," and "chronic disease management, screening everyone on admission, accreditation gave teeth to move these forward on the admin end," and "Increased timeline of how fast we are seeing people when they come in. Increasing the care they are receiving."[28]

Finally, several facilities commented on broader cultural change. One facility noted "the biggest change was the mentality of the whole staff in understanding why this process is important; took a lot of effort and heightening the understanding of staff; they must take steps in the standards for a reason" and another facility noted that they implemented "a lot more meetings, sitting down and collaborating."

Relationship between Custody and Medical Staff: About half of the treatment facilities noted substantial changes in the relationship between medical and custody. Medical leadership at one facility told us that "The staff had to have conversations about how processes will work, which meant discussing how custody versus medical execute certain processes." Medical staff reported that "Showing the deputies all of the different regulations and rules we need to follow that they need to help us with. Understand why we are asking for certain things. Really good education for us and custody. Makes it easier to do our work." As one facility leader noted: 'Yeah, when you look at officers they have a better understanding why they are making sure sick call slips are being given timely and being seen timely after submission rather than just giving aspirin."

IX Concluding Comments

Despite being the only U.S. population constitutionally guaranteed a right to adequate health care, incarcerated individuals face a dire situation in U.S. jails. In the last decade, official mortality rates within jails have increased sharply, particularly among the large unconvicted population. And our data suggest that the severity of the problem is substantially underreported, with mortality rates among our jails being three times higher than suggested by official statistics.

In this study, we evaluate whether accreditation, a commonplace method to assess quality in health care, can address the hidden health care crisis behind bars. Over a period of four years, we conducted the first RCT across jails and the first RCT of accreditation in health care in the U.S. Using original survey instruments developed by our research team, we find that NCCHC accreditation has statistically significant and economically meaningful effects on a range of outcomes. Accreditation improves measures of coordination among staff and improves compliance with quality standards, particularly those addressing personnel and training and patient care and treatment. We also find that accreditation substantially reduces mortality, with a reduction of up to 86% in the half-year following the expected time of the NCCHC on-site visit. These findings are supported by our qualitative evidence, which indicate that treatment facilities improved alignment between staff and implemented timely screenings of incarcerated individuals, particularly in the very first days of incarceration where mortality tends to be the highest. A back-of-the envelope benefit-cost calculation taking into account the reductions in mortality, recidivism, and litigation, as well as financial and personnel costs associated with accreditation, yields an annual net-benefit ranging from $2.7 to $56.4 million (see Appendix Section J).

Ultimately, our study shows that accreditation can improve health care and save lives in U.S. jails. There have been recent policy efforts to mandate accreditation across all jails (Berwick, Beckman and Gondi 2021). Our findings also illustrate that at least in the U.S. jail context, accreditation as a form of quality disclosure

[28] Dental care was also mentioned by several facilities, and it was noted to be particularly important given the lengthening stay of inmates in jails.

can inform staff of best practice, improve accountability and collaboration between agents, and yield positive and measurable effects on health outcomes.

Bibliography

Adler, Jessica L., and Weiwei Chen. 2023. "Jail Conditions And Mortality: Death Rates Associated With Turnover, Jail Size, And Population Characteristics." *Health Affairs*, 42(6): 849–857.

Alchian, Armen A, and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review*, 62(5): 777–795.

Alhawajreh, Mohammad J., Audrey S. Paterson, and William J. Jackson. 2023. "Impact of hospital accreditation on quality improvement in healthcare: A systematic review." *PLOS ONE*, 18: 1–20.

Alsan, Marcella, Arkey M. Barnett, Peter Hull, and Crystal Yang. 2024. ""Something Works" in U.S. Jails: Misconduct and Recidivism Effects of the IGNITE Program." *NBER Working Paper 32282*.

Alsan, Marcella, Crystal S. Yang, James R. Jolin, Lucy Tu, and Josiah D. Rich. 2023. "Health Care in U.S. Correctional Facilities—A Limited and Threatened Constitutional Right." *New England Journal of Medicine*, 388(9): 847–852.

American Medical Association. 1979. *American Medical Association Standards for Health Services in Jails.* American Medical Association Program to Improve Medical Care and Health Services in Correctional Institutions.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.

Association, American Bar Association & American Medical. 1974. *Medical and Health Care in Jails, Prisons, and Other Correctional Facilities: A Compilation of Standards and Materials.* ABA Commission on Correctional Facilities and Services and AMA Division of Medical Practice. Webb, Russell E.

Athey, Susan, and Guido W. Imbens. 2017. "Chapter 3 - The Econometrics of Randomized Experiments." In *Handbook of Economic Field Experiments*. Vol. 1, 73–140. Elsevier.

Bai, Jushan. 2003. "Inferential Theory for Factor Models of Large Dimensions." *Econometrica*, 71(1): 135–171.

Banerjee, Abhijit, Esther Duflo, Amy Finkelstein, Lawrence F Katz, Benjamin A Olken, and Anja Sautmann. 2020. "In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for RCTs in Economics." National Bureau of Economic Research Working Paper 26993.

Batistich, Mary Kate, William N. Evans, and David C. Phillips. 2021. "Reducing Re-arrests Through Light Touch Mental Health Outreach." Technical report, University of Notre Dame.

Bedard, Kelly, and Eric Helland. 2004. "The Location of Women's Prisons and the Deterrence Effect of "Harder" Time." *International Review of Law and Economics*, 24(2): 147–167.

Bedard, Kelly, and HE Frech III. 2009. "Prison Health Care: Is Contracting Out Healthy?" *Health Economics*, 18(11): 1248–1260.

Berwick, Donald M., Adam L. Beckman, and Suhas Gondi. 2021. "The Triple Aim Applied to Correctional Health Systems." *JAMA*, 325(10): 935–936.

Bhuller, Manudeep, Laura Khoury, and Katrine V Løken. n.d.. "Mental Health Consequences of Correctional Sentencing." *American Economic Journal: Economic Policy*.

Binswanger, Ingrid A., Marc F. Stern, Richard A. Deyo, Patrick J. Heagerty, Allen Cheadle, Joann G. Elmore, and Thomas D. Koepsell. 2007. "Release from Prison — A High Risk of Death for Former Inmates." *New England Journal of Medicine*, 356(2): 157–165.

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen. 2015. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." *The Review of Economic Studies*, 82(2): 457–489.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2020. "Healthy Business? Managerial Education and Management in Health Care." *The Review of Economics and Statistics*, 102(3): 506–517.

Bogh, Søren Bie, Anne Mette Falstie-Jensen, Erik Hollnagel, René Holst, Jeffrey Braithwaite, and Søren Paaske Johnsen. 2017. "Improvement in quality of hospital care during accreditation: A nationwide stepped-wedge study." *International Journal for Quality in Health Care*, 28(6): 715–720.

Bovell-Ammon, Benjamin J., Ziming Xuan, Michael K. Paasche-Orlow, and Marc R. LaRochelle. 2021. "Association of Incarceration with Mortality by Race from a National Longitudinal Cohort Study." *JAMA Network Open*, 4(12): e2133083–e2133083.

Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics*, 1(4): 200–232.

Carolynn, Steinwald, and Alevizos Gus. 1973. "Medical Care in U.S. Jails: A 1972 AMA Survey." American Medical Association. *US Department of Justice, Office of Justice Programs.*

Carson, E. Ann. 2021. "Mortality in Local Jails, 2000–2019-Statistical Tables." Mortality in Local Jails and State Prisons. *US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics*.

Chassin, Mark R. 2021. "Personal Communication." August, 2021. Notes available from https://www.dropbox.com/scl/fi/bb48ojpxcv23b2ht69p8q/Joint-Commission-Dr-Chassin_2aug2021.docx?rlkey=nffjk24l2yinugsjt5wqdam95&dl=0.

Chen, Min. 2008. "Minimum Quality Standards and Strategic Vertical Differentiation: An Empirical Study of Nursing Homes." *Northwestern University PhD Dissertation*.

Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor. 2010. "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." *Journal of Human Resources*, 45(3): 655–681.

Cloud, David H., Brie Williams, Regine Haardörfer, Lauren Brinkley-Rubinstein, and Hannah L.F. Cooper. 2023. "Extreme Heat and Suicide Watch Incidents Among Incarcerated Men." *JAMA Network Open*, 6(8): e2328380–e2328380.

Coll, Steve. 2019. "The Jail Health-Care Crisis." *The New Yorker*. [Internet] Accessed November, 2024. Available from https://www.newyorker.com/magazine/2019/03/04/the-jail-health-care-crisis.

Congressional Research Service. 2023. "Death in Custody Reporting Act: Background and Legislative Considerations." *Congressional Research Service*

Cowan, Kristen N., Katherine Lemasters, and Lauren Brinkley Rubinstein. 2024. "Temperatures Extremes and Mortality During Summer Months in Federal Prisons Across the United States, 2010-2021." Vol. 2024.

Crémer, Jacques, and Richard P McLean. 1988. "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions." *Econometrica*, 1247–1257.

Curran, Jill, Brendan Saloner, Tyle Winkelman, and G. Caleb Alexander. 2023. "Estimated Use of Prescription Medications Among Individuals Incarcerated in Jails and State Prisons in the U.S." *JAMA Health Forum*, 4(4): e230482–e230482.

Daza, Sebastian, Alberto Palloni, and Jerrett Jones. 2020. "The Consequences of Incarceration for Mortality in the United States." *Demography*, 57(2): 577–598.

Doleac, Jennifer L. 2023. "Encouraging Desistance from Crime." *Journal of Economic Literature*, 61(2): 383–427.

DOT. 2024. "Departmental Guidance on Valuation of a Statistical Life in Economic Analysis." [Internet] Accessed December, 2024. Available from https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-on-valuation-of-a-statistical-life-in-economic-analysis.

Douds, Anne S., and Eileen M. Ahlin. 2016. "Do NCCHC Dental Standards Have Any Teeth?" *Journal of Correctional Health Care*, 22(3): 180–188.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature*, 48(4): 935–963.

Durose, Matthew R., and Leonardo Antenangeli. 2021. "Recidivism of Prisoners Released in 34 States in 2012: A 5-year Follow-Up Period (2012-2017)." Recidivism of Prisoners Released in 34 States in 2012: A 5-year Follow-Up Period (2012-2017). *US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics*.

Edmonds, Mira. 2020. "The Reincorporation of Prisoners into the Body Politic: Eliminating the Medicaid Inmate Exclusion Policy." *Georgetown Journal on Poverty Law and Policy*, 28: 279.

El-Sabawi, Taleed, Shelly R Weizman, Somer M Brown, and Regina M LaBelle. 2023. "Dying Inside: Litigation Patterns for Deaths in Jail Custody." *Journal of Correctional Health Care*, 29(4): 275–281.

Estelle v. Gamble. 1976. "Estelle v. Gamble." 429 U.S. 97.

Fair, Helen, and Roy Walmsley. 2024. "World Prison Population List." Institute for Crime & Justice Policy Research.

Fiscella, Kevin, Leo Beletsky, and Sarah E. Wakeman. 2017. "The Inmate Exception and Reform of Correctional Health Care." *American Journal of Public Health*, 107(3): 384–385.

GAO. 2022. "Deaths in Custody: Additional Action Needed to Help Ensure Data Collected by DOJ Are Utilized." *GAO*, 2022.

Gibson, Brent R., and Gary Phillips. 2016. "Challenges and Opportunities in Correctional Health Care Quality: a Descriptive Analysis of Compliance with NCCHC Standards." *Journal of Correctional Health Care*, 22(4): 280–289.

Goldhaber, Dan, and Emily Anthony. 2007. "Can teacher quality be effectively assessed? National board certification as a signal of effective teaching." *The Review of Economics and Statistics*, 89(1): 134–150.

Greene, Meredith, Cyrus Ahalt, Irena Stijacic-Cenzer, Lia Metzger, and Brie Williams. 2018. "Older Adults in Jail: High Rates and Early Onset of Geriatric Conditions." *Health & justice*, 6: 1–9.

Greenfield, David, and Jeffrey Braithwaite. 2008. "Health Sector Accreditation Research: A Systematic Review." *International Journal for Quality in Health Care*, 20(3): 172–183.

Greifinger, Robert B. 2007. "Thirty Years Since Estelle v. Gamble: Looking Forward, Not Wayward." In *Public Health Behind Bars: From Prisons to Communities*. 1–10. Springer.

Hardin, James W., Henrik Schmiediche, and Raymond J. Carroll. 2003. "Instrumental Variables, Bootstrapping, and Generalized Linear Models." *The Stata Journal*, 3(4): 351–360.

Health Forum. 2024. "AHA Hospital Statistics." 2024 Edition. *AHA Data & Insights. American Hospital Association.*

Henrichson, Christian, Joshua Rinaldi, and Ruth Delaney. 2015. "The Price of Jails: Measuring the Taxpayer Cost of Local Incarceration." *Vera Institute of Justice*.

Hjalmarsson, Randi, and Matthew J. Lindquist. 2022. "The Health Effects of Prison." *American Economic Journal: Applied Economics*, 14(4): 234–270.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *The Bell Journal of Economics*, 324–340.

Howell, Benjamin A., Laura C. Hawks, Lilanthi Balasuriya, Virginia W. Chang, Emily A. Wang, and Tyler Winkelman. 2023. "Health Insurance and Mental Health Treatment Use Among Adults with Criminal Legal Involvement After Medicaid Expansion." *Psychiatric Services*, 74(10): 1019–1026.

Jha, Ashish K. 2018. "Accreditation, Quality, and Making Hospital Care Better." *JAMA*, 320(23): 2410–2411.

Jin, Ginger Zhe, and Phillip Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *Quarterly Journal of Economics*, 118(2): 409–451.

Johnson, Rucker C., and Steven Raphael. 2009. "The Effects of Male Incarceration Dynamics on Acquired Immune Deficiency Syndrome Infection Rates Among African American Women and Men." *The Journal of Law and Economics*, 52(2): 251–293.

Kane, Thomas J, Jonah E Rockoff, and Douglas O Staiger. 2008. "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education review*, 27(6): 615–631.

Katz, Lawrence, Steven D. Levitt, and Ellen Shustorovich. 2003. "Prison Conditions, Capital Punishment, and Deterrence." *American Law and Economics Review*, 5(2): 318–343.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.

Lise, Jeremy, and Fabien Postel-Vinay. 2020. "Multidimensional Skills, Sorting, and Human Capital Accumulation." *American Economic Review*, 110(8): 2328–2376.

Maner, Morgan, Marisa Omori, Lauren Brinkley-Rubinstein, Curt G. Beckwith, and Kathryn Nowotny. 2022. "Infectious Disease Surveillance in U.S. Jails: Findings from a National Survey." *PLoS One*, 17(8): e0272374.

McCann, Sam. 2024. "The Hidden Deaths in American Jails." *Vera*, 2024. [Internet] Accessed November, 2024. Available from https://www.vera.org/news/the-hidden-deaths-in-american-jails#:~:text=They%20exist%20for%20one%20simple,death%20data%20is%20deliberately%20opaque.

McLeod, Marsha. 2019. "The Private Option." *The Atlantic*. [Internet] Accessed November, 2024. Available from https://www.theatlantic.com/politics/archive/2019/09/private-equitys-grip-on-jail-health-care/597871/.

Miller, Ted R, Mark A Cohen, David I Swedler, Bina Ali, and Delia V Hendrie. 2021. "Incidence and Costs of Personal and Property Crimes in the USA, 2017." *Journal of Benefit-Cost Analysis*, 12(1): 24–54.

Mitchell Jr, Roger A, and Jay D Aronson. 2023. *Death in Custody: How America Ignores the Truth and What We Can Do About It*. Johns Hopkins University Press.

Modlin, Herbert C. 1979. "Medical Care in Correctional Institutions: The AMA project." *Bulletin of the American Academy of Psychiatry and the Law*, 7: 118–124.

Mukherjee, Anita, and Nicholas J Sanders. 2021. "The Causal Effect of Heat on Violence: Social Implications of Unmitigated Heat Among the Incarcerated." National Bureau of Economic Research Working Paper 28987.

Mumford, Virginia, Kevin Forde, David Greenfield, Reece Hinchcliff, and Jeffrey Braithwaite. 2013. "Health Services Accreditation: What is the Evidence that the Benefits Justify the Costs?" *International Journal for Quality in Health Care*, 25(5): 606–620.

NASEM. 2019. *The Effects of Incarceration and Reentry on Community Health and Well-Being: Proceedings of a Workshop.* National Academies Press (US). National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Population Health and Public Health Practice; Roundtable on the Promotion of Health Equity. Karen M. Anderson and Steve Olson, editors.

National Commission on Correctional Health Care. 2018. "NCCHC Historical Time Line." [Internet] Accessed June, 2024. Available from https://www.ncchc.org/historical-time-line/.

National Commission on Correctional Health Care. 2021*a*. "The Importance of Custody-Health Care Collaboration, Part 1." [Internet] Accessed November, 2024. Available from https://www.ncchc.org/the-importance-of-custody-health-care-collaboration-part-1/.

National Commission on Correctional Health Care. 2021*b*. "The Importance of Custody-Health Care Collaboration, Part 2." [Internet] Accessed November, 2024. Available from https://www.ncchc.org/the-importance-of-custody-health-care-collaboration-part-2/.

National Sheriffs' Assocation. 2015. "Office of Sheriff State-By-State Elections Information."

Neisser, Eric. 1977. "Is There a Doctor in the Joint the Search for Constitutional Standards for Prison Health Care?" *Virginia Law Review*, 63: 921.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver. 2024. "The Effect of Incarceration on Mortality." *Review of Economics and Statistics*, 106(4): 956–973.

Patterson, Evelyn J. 2010. "Incarcerating Death: Mortality in U.S. State Correctional Facilities, 1985–1998." *Demography*, 47: 587–607.

Pew Charitable Trusts. 2014. "Managing Prison Health Care Spending." Pew Charitable Trusts and John D. and Catherine T. MacArthur Foundation.

Pew Charitable Trusts. 2017. "Prison Health Care: Costs and Quality."

Prout, Curtis, and Robert N. Ross. 1988. *Care and Punishment: The Dilemmas of Prison Medicine.* University of Pittsburgh Press Pittsburgh, PA.

Puglisi, Lisa B., and Emily A. Wang. 2021. "Health Care for People Who are Incarcerated." *Nature Reviews Disease Primers*, 7(1): 50.

Reaves, Brian, and Matthew J. Hickman. 2002. *Census of State and Local Law Enforcement Agencies, 2000.* U.S. Department of Justice, Office of Justice Programs.

Reinhart, Eric, and Daniel L. Chen. 2021. "Carceral-Community Epidemiology, Structural Racism, and COVID-19 Disparities." *Proceedings of the National Academy of Sciences*, 118(21): e2026577118.

Rich, Josiah D., Scott A. Allen, and Brie A. Williams. 2015. "The Need For Higher Standards in Correctional Healthcare to Improve Public Health." *Journal of General Internal Medicine*, 30: 503–507.

Ruch, Donna A., Danielle L. Steelesmith, Guy Brock, Samantha J. Boch, Camille R. Quinn, Jeffrey A. Bridge, John V. Campo, and Cynthia A. Fontanella. 2021. "Mortality and Cause of Death Among Youths Previously Incarcerated in the Juvenile Legal System." *JAMA network open*, 4(12): e2140352–e2140352.

Safety Source. 2021. "National Directory of Law Enforcement Administrators."

Salmon, Jack Warren, John Heavens, Carl F Lombard, and Paula Tavrow. 2003. "The Impact of Accreditation on the Quality of Hospital Care: KwaZulu-Natal Province, Republic of South Africa." *Quality Assurance Project*.

Saloner, Brendan. 2023. "A Chance to Modernize Health Care Behind Bars-Section 1115 Medicaid Inmate Exclusion Waivers." *The New England Journal of Medicine*, 389(16): 1449–1451.

Sechrest, Dale K. 1976. "The Accreditation Movement in Corrections." *Federal Probation*, 40: 15. *U.S. Department of Justice. National Institute of Justice. Office of Research Programs.*

Shaw, Charles D. 2003. "Evaluating Accreditation." *International Journal for Quality in Health Care*, 15(6): 455–456.

Skarha, Julianne, Amite Dominick, Keith Spangler, David Dosa, Josiah D. Rich, David A. Savitz, and Antonella Zanobetti. 2022. "Provision of Air Conditioning and Heat-Related Mortality in Texas Prisons." *JAMA Network Open*, 5(11): e2239849–e2239849.

Skarha, Julianne, Keith Spangler, David Dosa, Josiah D. Rich, David A. Savitz, and Antonella Zanobetti. 2023. "Heat-Related Mortality in U.S. State and Private Prisons: A Case-Crossover Analysis." *PLoS One*, 18(3): e0281389.

Smith, Amy. 2013. *Health and Incarceration: A Workshop Summary.* National Academies Press. National Research Council; Division of Behavioral and Social Sciences and Education; Institute of Medicine; Committee on Law and Justice; Board on the Health of Select Populations; Committee on Causes and Consequences of High Rates of Incarceration.

Social Security Amendments. 1965. "Social Security Amendments (Medicare and Medicaid Act)."

Spaulding, Anne C., Ryan M. Seals, Victoria A. McCallum, Sebastian D. Perez, Amanda K. Brzozowski, and N. Kyle Steenland. 2011. "Prisoner Survival Inside and Outside of the Institution: Implications for Health-Care Planning." *American Journal of Epidemiology*, 173(5): 479–487.

Stanley, Judith A. 2009. "A Personal Retrospective: In the Eye of the Accreditation Storm (Part I of II)." *Journal of Correctional Health Care*, 15(2): 150–156.

Strausz, Roland. 2005. "Honest Certification and the Threat of Capture." *International Journal of Industrial Organization*, 23(1-2): 45–62.

The Joint Commission. 2018. "Safety Culture Assessment Improving the survey process." 28(6).

The Joint Commission. 2021. "Value of Accreditation and Certification Literature Database." [Internet] Accessed November, 2024. Available from https://manual.jointcommission.org/Accreditation/WebHome.

The Joint Commission. 2024. "Hospital Accreditation Fact Sheet." [Internet] Accessed November, 2024. Available from https://www.jointcommission.org/resources/news-and-multimedia/fact-sheets/facts-about-hospital-accreditation/.

Tuholske, Cascade, Victoria D. Lynch, Raenita Spriggs, Yoonjung Ahn, Colin Raymond, Anne E. Nigra, and Robbie M. Parks. 2024. "Hazardous Heat Exposure Among Incarcerated People in the United States." *Nature Sustainability*, 7(4): 394–398.

U.S. Department of Justice. 2022. "Census of Jails, 2019." United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. Inter-university Consortium for Political and Social Research [distributor]. Accessed November, 2024. Data available from https://www.icpsr.umich.edu/web/NACJD/studies/38323.

Wang, Leah. 2021. "Rise in Jail Deaths is Especially Troubling as Jail Populations Become More Rural and More Female." *Prison Policy Initiative*, 2021. [Internet] Accessed November, 2024. Available from https://www.prisonpolicy.org/blog/2021/06/23/jail_mortality/#:~:text=New%20data%20show%20record%20high,ensures%20more%20deaths%20to%20come.

Wildeman, Christopher, E. Ann Carson, Daniela Golinelli, Margaret E. Noonan, and Natalia Emanuel. 2016. "Mortality Among White, Black, and Hispanic Male and Female State Prisoners, 2001–2009." *SSM-Population Health*, 2: 10–13.

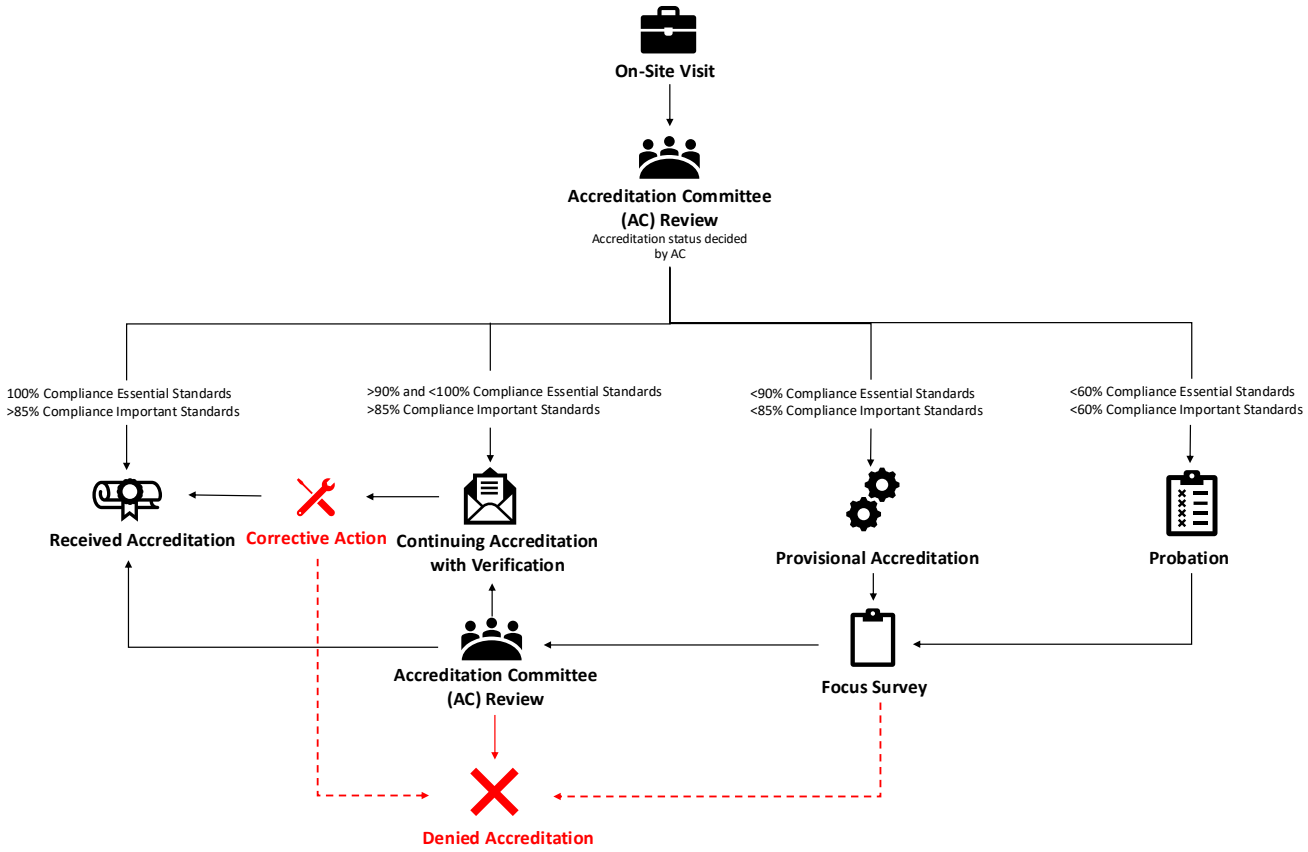
Williams, Brie, Michele DiTomas, and Alison Pachynski. 2021. "The Growing Geriatric Prison Population: A Dire Public Health Consequence of Mass Incarceration." *Journal of the American Geriatrics Society*, 69(12): 3407–3409.

Wishart, Margaret D., and Nancy N. Dubler. 1983. *Health Care in Prisons, Jails and Detention Centers: Some Legal and Ethical Dilemmas.* New York: Montefiore Medical Center.

Zeng, Zhen. 2023. "Jail Inmates in 2022 – Statistical Tables." U.S. Department of Justice, Bureau of Justice Statistics.

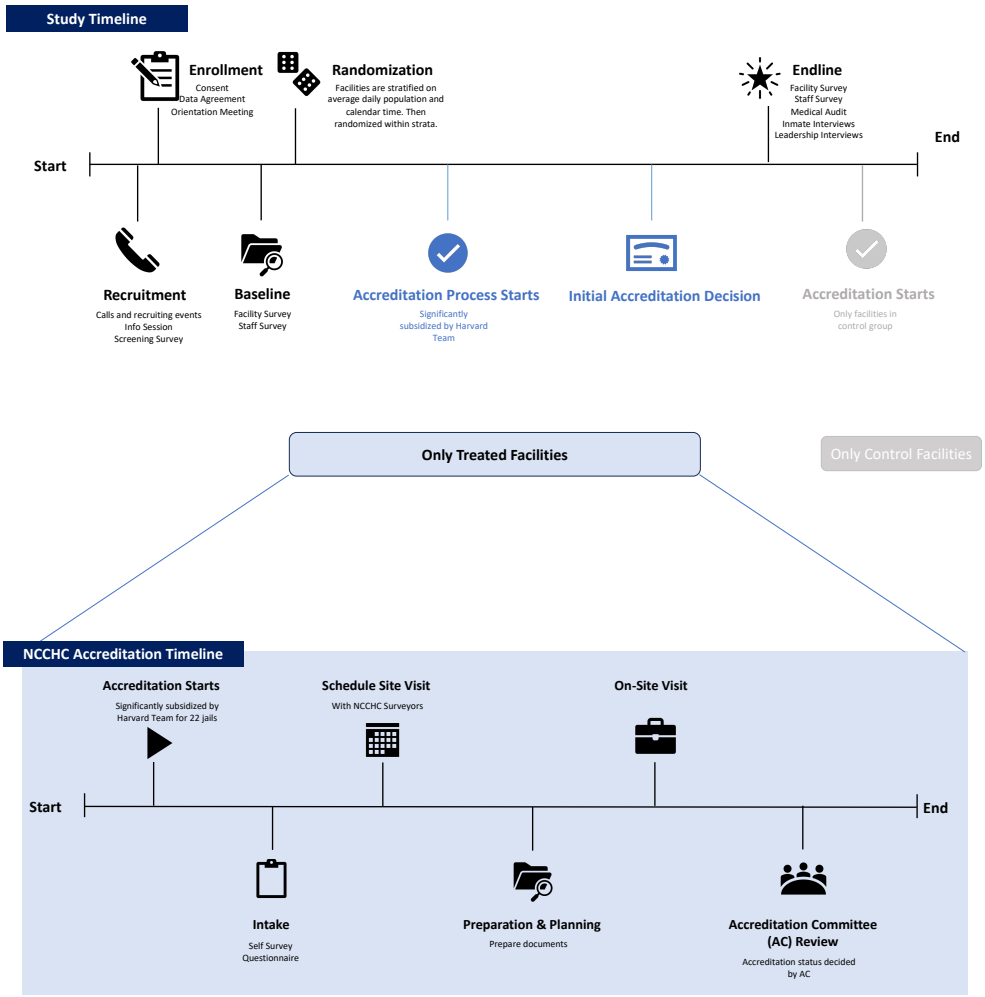
Zeng, Zhen, and Todd D. Minton. 2021. "Census of jails, 2005-2019-Statistical Tables." *U.S. Department of Justice, Bureau of Justice Statistics*, 255406: 1–51.

Figure 1: Accreditation Decision Process



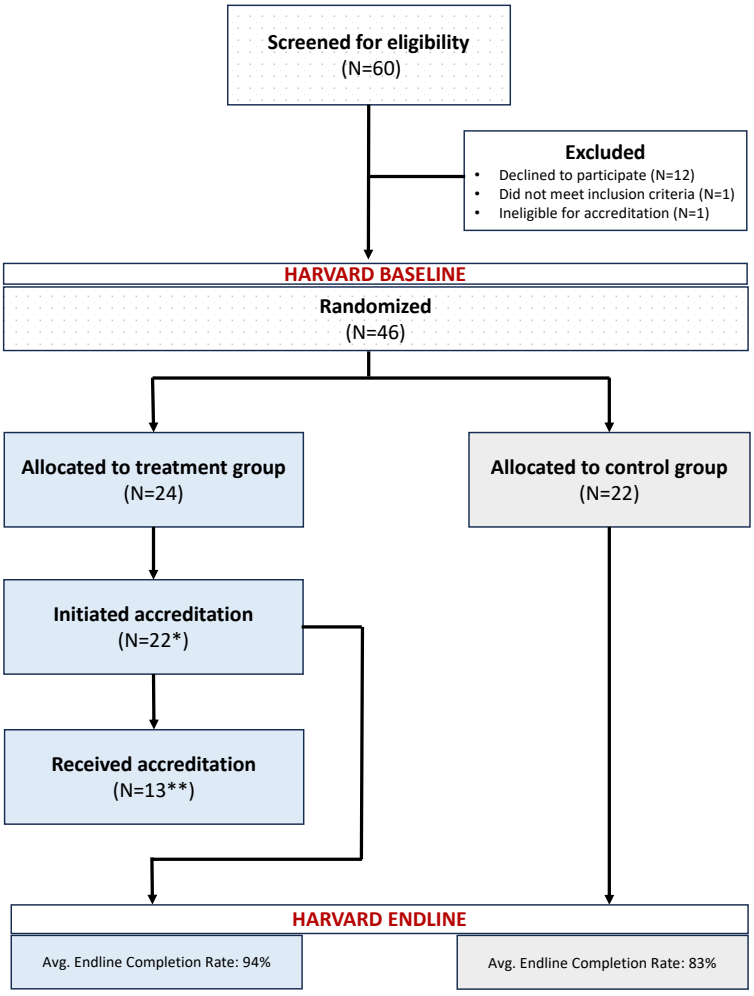
Notes: This figure depicts the NCCHC accreditation process. The holistic process is based in part on compliance with important and essential standards. Facilities that comply with less than 90% of essential standards and less than 85% of important standards have apparent deficiencies that significantly lower their compliance levels below that acceptable for accreditation and are placed on "Provisional Accreditation" status. Facilities that comply with less than 60% of essential and important standards have excessive standard noncompliance and are placed on "Probation" status.

Figure 2: Study Timeline



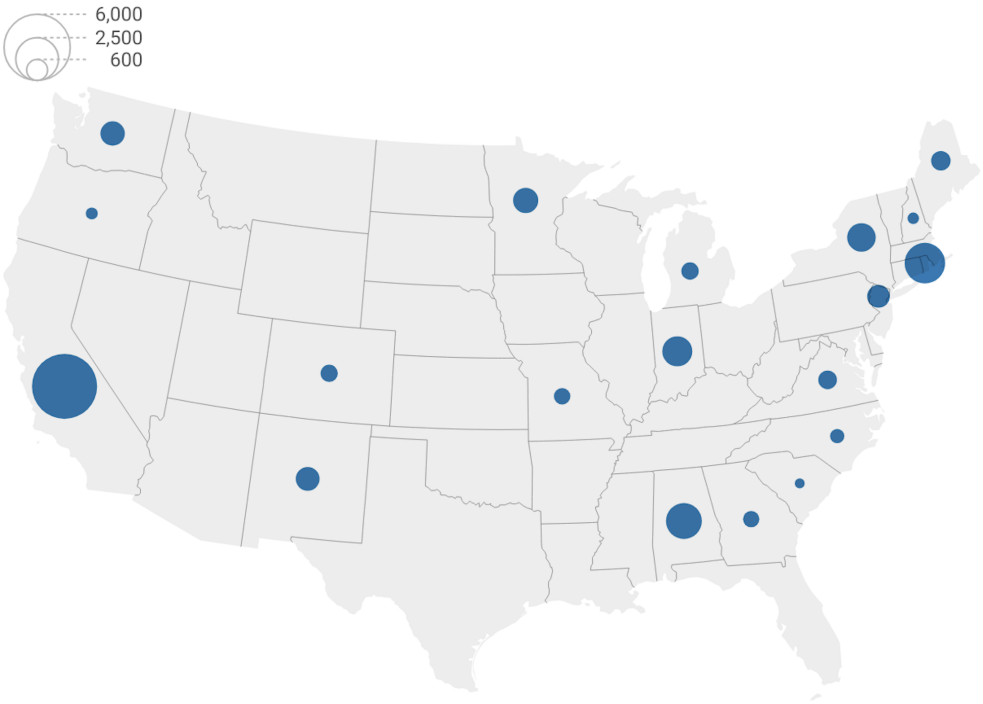
Notes: This figure depicts the Harvard study timeline. The top panel illustrates the Harvard study timeline from recruitment to endline for all facilities. The bottom panel illustrates the NCCHC accreditation process for facilities randomized to the treatment group who initiated the accreditation process. See also Appendix Section C for additional details.

Figure 3: Study Flow Overview



Notes: This figure provides an overview of the Harvard study flow. 60 facilities were screened for eligiblity for the study. Of those 60, 12 did not want to participate, one did not qualify due to not meeting the inclusion criteria, and one was ineligible to participate in accreditation due to the structure of health services at the facility. The Harvard study baseline completion rate is 100% as it was a prerequisite for randomization. The average endline completion rate is computed as the proportion of facilities that completed the endline facility survey, endline staff survey, medical audit, and provided a death log. Therefore, the denominator of the endline completion rate for both the treatment and control group is 4x22=88. Out of the 24 facilities allocated to treatment, 2 never initiated accreditation for reasons orthogonal to assignment, as discussed in the main text, and are excluded from the main analytical sample. These two facilities had average endline completion rates that are much lower (25%) than the other facilities. Results remain largely unchanged when including these two facilities, using baseline to impute missing values (see Appendix Table A17 and A18). * 2 facilities did not begin accreditation; ** 1 facility is still under review but is very close to receipt of accreditation. The draft will be updated upon its completion.

Figure 4: Map of Study Jails and ADP



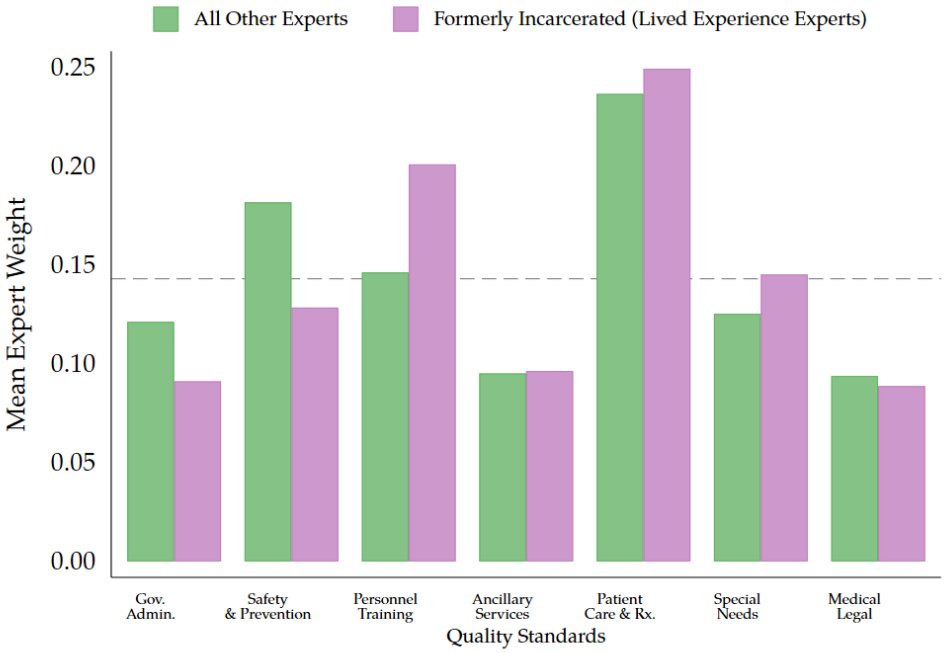
Notes: This map depicts the geographic location of the facilities in our main analytical sample (N=44). The size of the circle represents the total ADP across facilities within that state.

Figure 5: Example from Facility Survey Coding Manual

Table of Contents								
Variable Name	**Reason for Potential Skip**	**Format**	**Value if no Information**	**Question**	**Coder to Determine Instructions**	**Valid Entries**	**Variable Label**	**Variable Description**
upl_griev_pol		*Text*		*Please upload the protocol for addressing health-related grievances*	***Filled by facility***		*Grievance Policy*	*Indicates status of facility uploaded document*
A10_GRV01_F	Skip if answer to upl_griev_pol is not "Document Uploaded" or "In policy and procedures manual"	Yes (1) - No (0)		Does the documentation uploaded by the facility include a protocol for addressing health-related grievances?	Coder to look at uploaded documentation/facility answer (Based on document upload - upl_griev_pol)	0 and 1	Grievance Policy: Check	Indicates whether there is a grievance policy for health care complaints
A10_GRV02_F	Skip if A10_GRV01_F was skipped, or the answer is No (0)	Integer	-99	Based on the grievance policy, once someone files a grievance, how many days until there is a formal response?	Determine Answer by Checking Grievance Policy. A simple verbal acknowledgement of the grievance does not count as a formal response (Based on document upload - upl_griev_pol)	Integers	Grievances: Days for Formal Response	States, based on grievance policy, how many days until there is a formal response after someone files a grievance
A10_GRV03_F	Skip if A10_GRV01_F was skipped, or the answer is No (0)	Yes (1) - No (0)		Does the grievance policy indicate that there is a process for appealing?	Determine Answer by Checking Grievance Policy (Based on document upload - upl_griev_pol)	0 and 1	Grievances: Appeal Process	Indicates, based on grievance policy, whether there is a process for appealing
A10_GRV04_F	Skip if A10_GRV01_F was skipped, or the answer is No (0)	Yes (1) - No (0)		Does the grievance policy indicate that there are formal records of health-related grievances?	Coder to look at uploaded documentation/facility answer (Based on document upload - upl_griev_pol)	0 and 1	Grievances: Formal Records	Indicates whether there are formal records of health-related grievances from the past month

Notes: This figure depicts an example from our coding manual that assists coders in determining whether a facility complies with standards on grievance policies.

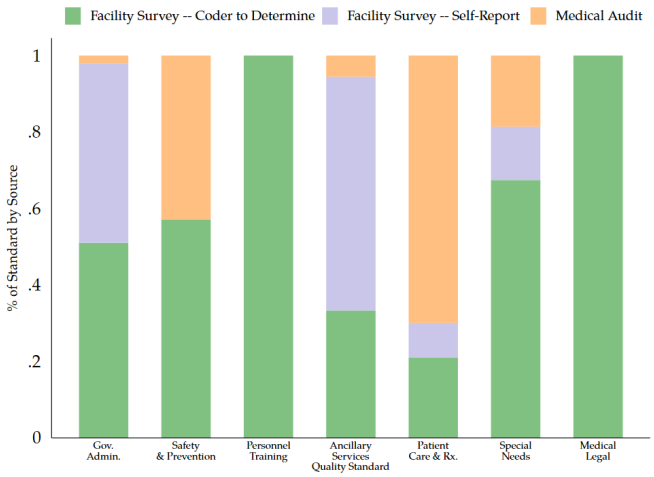
Figure 6: Expert Weights



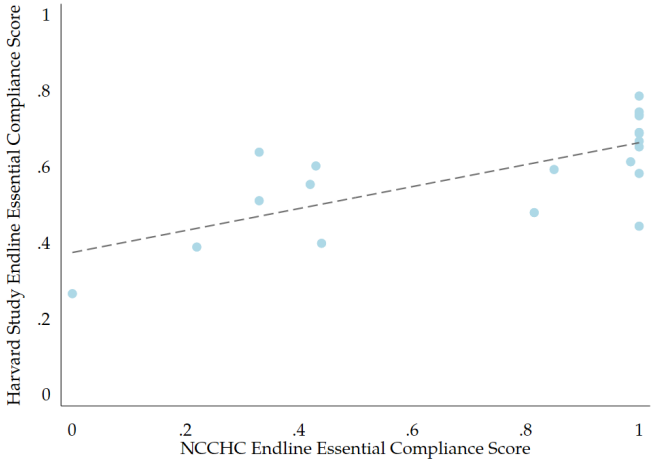
Notes: This figure shows the reciprocal rank weights placed on the seven quality standards by the formerly incarcerated (lived experience experts) and other experts (jail administrators, jail health professionals, and policy and public health experts).

Figure 7: Measurement of Quality Standards

Panel (A) Composition of Standards

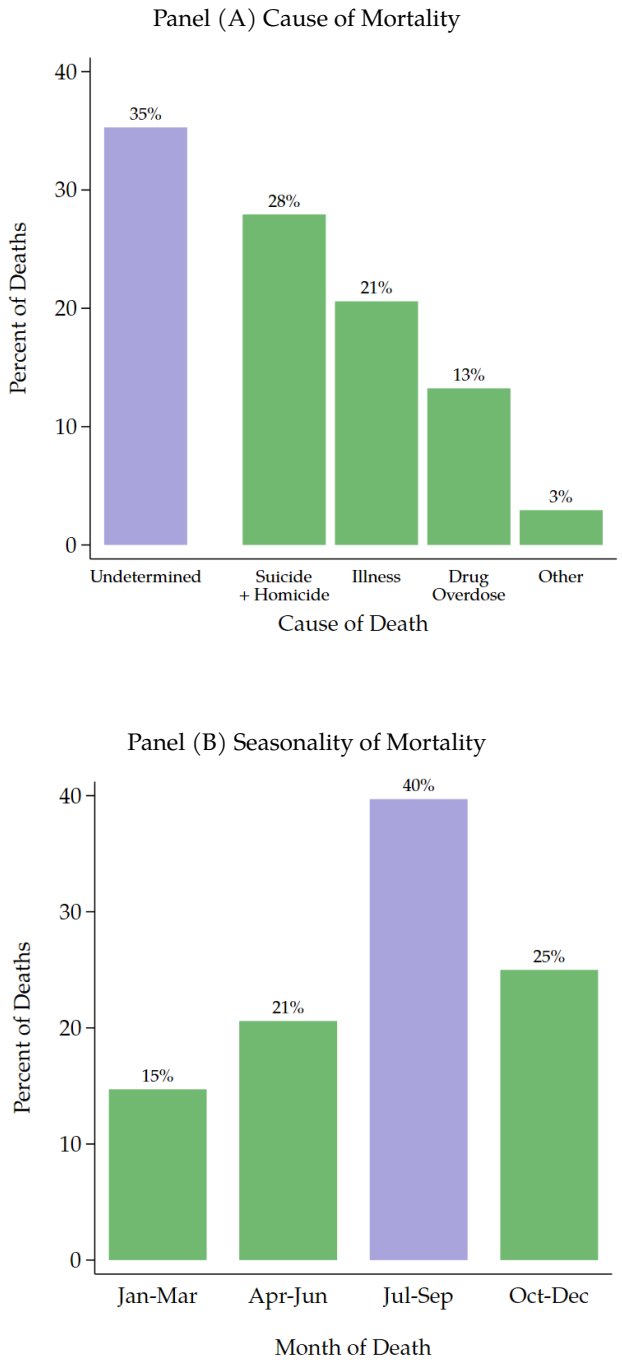


Panel (B) Corr. Harvard and NCCHC Compliance



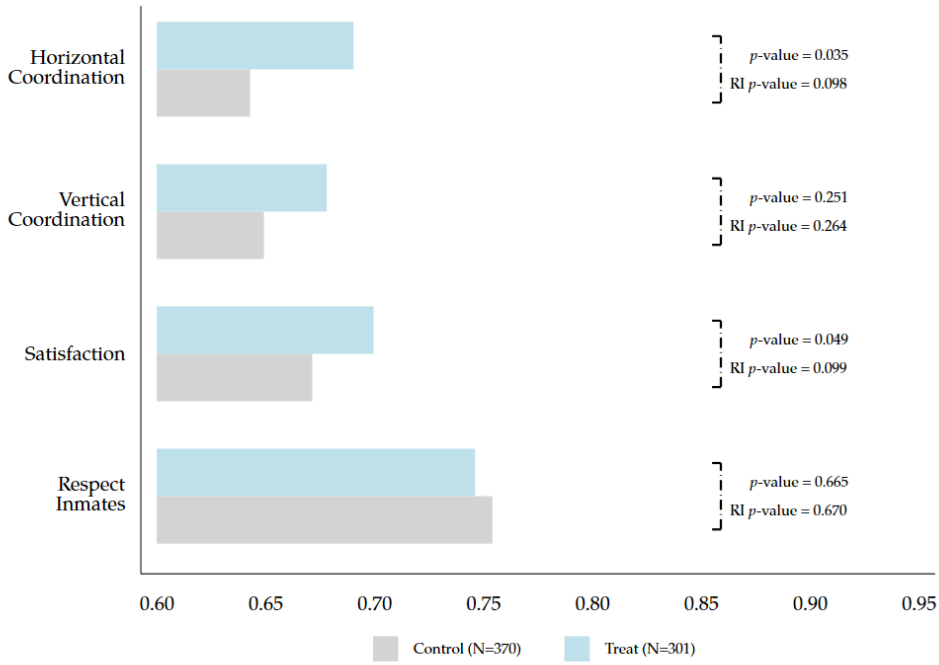
Notes: Panel (A) shows the proportion of variables used to construct the seven quality standards from the (i) facility survey – self-reported, (ii) the facility survey – coder to decide, and (iii) the medical audits. Panel (B) shows the correlation between the Harvard study compliance scores and the NCCHC compliance scores. The study score follows the same approach as used in the NCCHC accreditation process, but uses the study-collected data instead.

Figure 8: Distribution of Mortality by Cause and Season



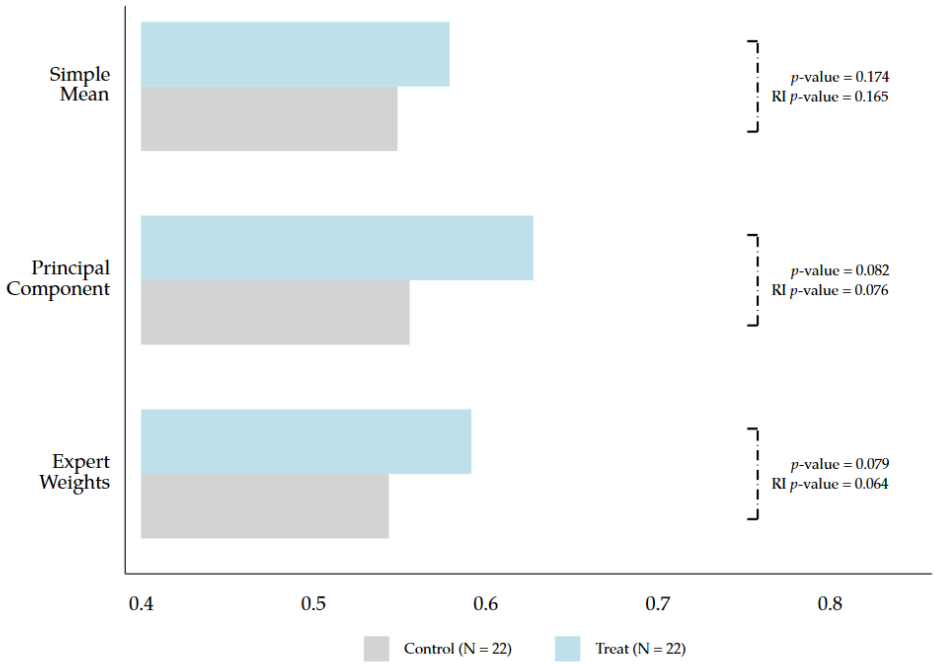
Notes: This figure shows cause and month of death using data from death logs provided by jails. Four facilities in the analytical sample did not provide a death log so media reported deaths are used instead. Panel (A) disaggregates the total number of deaths in the sample by cause of death and plots the share of each cause as a percentage of total deaths. If the cause is unknown, undetermined, or if the report is pending, it is classified as undetermined. In the event of a death by suicide or homicide, it is classified as such. If a pre-existing condition, infection, or any other physical ailment is indicated as the cause of death, the death falls under illness. In the event of a drug toxicity, or overdose, the death is labeled as a drug overdose. If a death does not fall into any of these four categories, it is classified as other. Panel (B) shows the seasonality of deaths by plotting the mean number of deaths for each quarter of the year. The date of death indicated for each death in the death log is used to identify the quarter of the year.

Figure 9: Effect of Accreditation Assignment on Staff Sentiment Indices: Reduced Form Estimates



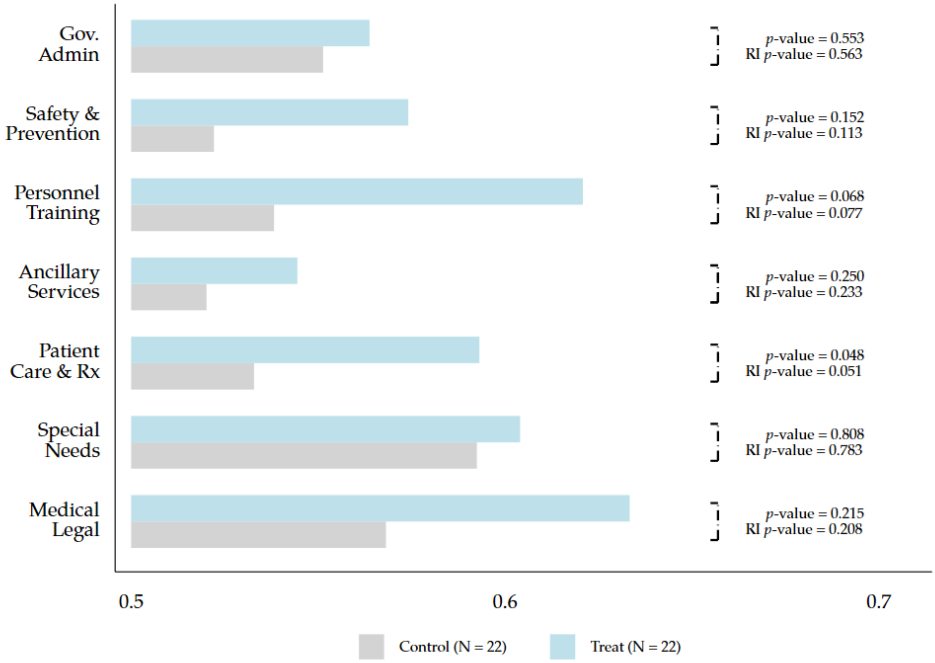
Notes: This figure shows the treatment effect of accreditation assignment on four staff sentiment indices. Staff sentiment indices are created using data from the baseline/endline staff survey. The horizontal coordination index (bar 1) includes: (1) "Custody staff support the implementation of clinical decisions and partner with clinical staff"; (2) "Custody and clinical team staff work collaboratively to give good health care to inmates"; (3) "There is open communication between custody and health staff." The vertical coordination index (bar 2) includes: (1) "Staff feel comfortable speaking up when they see something that may negatively affect inmate care"; (2) "When staff speak up about inmate safety and care decisions, those with greater authority are open to their concerns"; (3) "My supervisor considers staff suggestions for improving inmate health and safety"; (4) "My supervisor considers staff suggestions for improving staff health and safety"; (5) "My supervisor takes direct actions to address staff concerns for improving inmate health and safety"; (6) "My supervisor takes direct actions to address staff concerns for improving staff health and safety." Questions (1) and (2) are only asked to health staff. The satisfaction index (bar 3) includes: (1) "I received the training I need to do my job well"; (2) "I have the tools and resources I need to do my job well"; (3) "I am satisfied with my job"; (4) "I find my job to be meaningful"; (5) "I feel valued at my job"; (6) "I find my job to be challenging"; (7) "I would recommend my job to others." The respect inmates index (bar 4) includes: (1) "Inmates at this facility are treated with respect." These questions are answered on a Likert scale and renormalized on a scale from 0 to 1 with higher values representing stronger agreement. In all bars, the control mean (gray) is calculated as the mean of staff in the control group. Treatment effects are obtained from estimating Equation 1, controlling for the baseline mean of the respective outcome variable at the facility level and randomization strata fixed effects. The treatment mean (blue) plotted is calculated as the control mean plus the estimated treatment effect. We obtained endline responses from 40 jails and we impute missing endline responses for the other 4 facilities in our sample with baseline values. p-values based on robust standard errors and randomization inference (RI) are shown.

Figure 10: Effect of Accreditation Assignment on Meta Quality Standards: Reduced Form Estimates



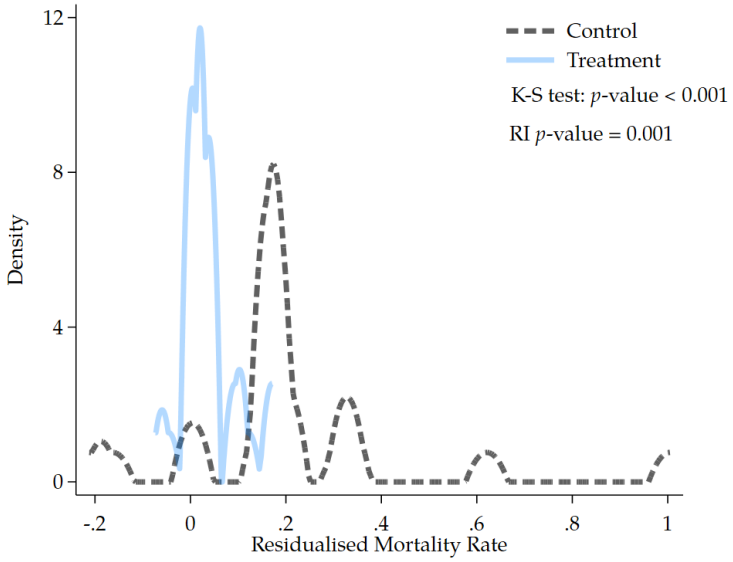
Notes: This figure shows the treatment effect of accreditation assignment on meta quality standards constructed by combining the seven quality standards in three ways as described in the main text: equal weights, the first component using principal component analysis, and reciprocal rank weights using our expert survey responses. For detailed definitions of the standards and their calculation see Appendix Section I and for expert weights see Figure 6. In all bars, the control mean (gray) is calculated as the mean of facilities in the control group. Treatment effects are obtained from estimating Equation 1, controlling for the baseline mean of the respective outcome variable at the facility level and randomization strata fixed effects. The treatment mean (blue) plotted is calculated as the control mean plus the estimated treatment effect. *p*-values based on robust standard errors and randomization inference (RI) are shown.

Figure 11: Effect of Accreditation Assignment on Quality Standards: Reduced Form Estimates



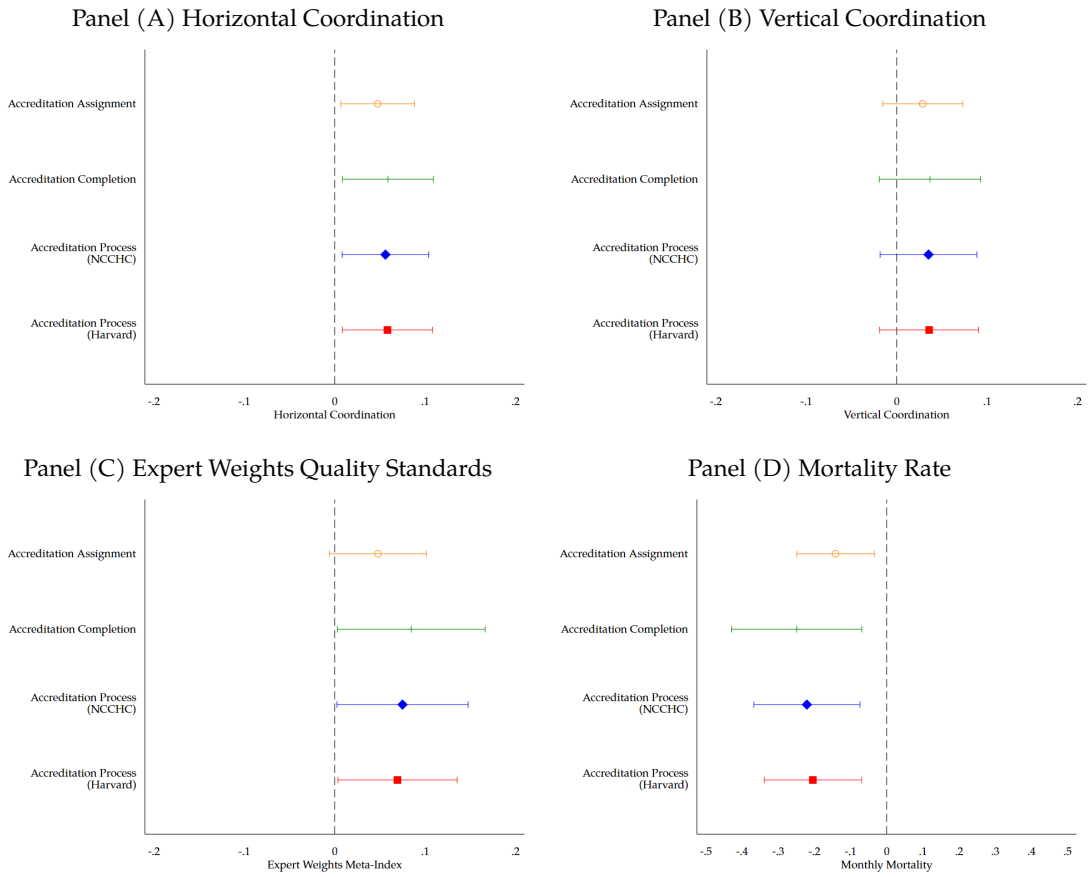
Notes: This figure shows the treatment effect of accreditation assignment on the seven quality standards. Quality standards are created using data from the baseline/endline facility survey and look-back medical audit. For detailed definitions of the standards and their calculation, see Appendix Section I). In all bars, the control mean (gray) is calculated as the mean of facilities in the control group. Treatment effects are obtained from estimating Equation 1, controlling for the baseline mean of the respective outcome variable at the facility level and randomization strata fixed effects. The treatment mean (blue) plotted is calculated as the control mean plus the estimated treatment effect. *p*-values based on robust standard errors and randomization inference (RI) are shown.

Figure 12: Effect of Accreditation Assignment on Mortality: Reduced Form Estimates



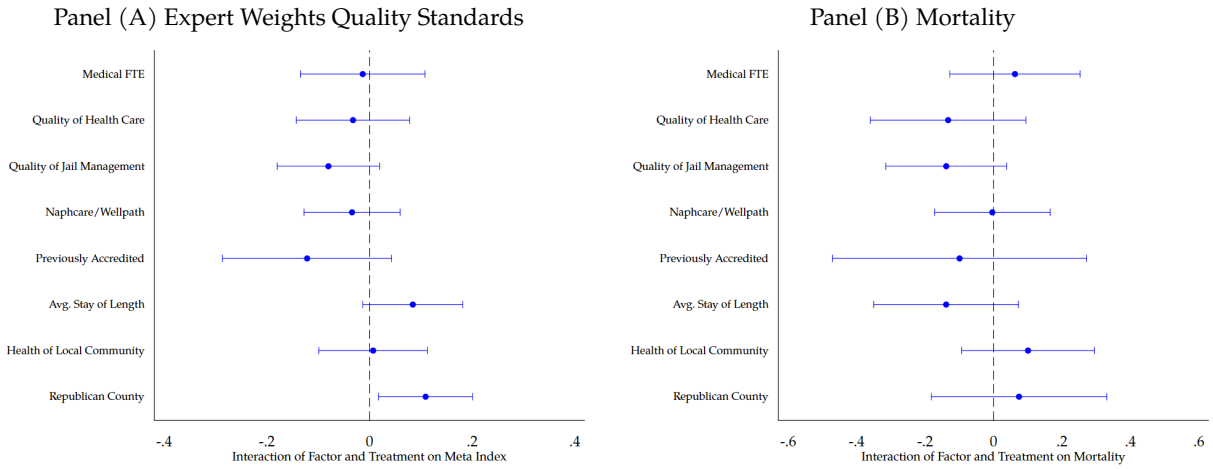
Notes: This figure shows the distribution of mortality for the treatment and control facilities. The mean monthly mortality rate is residualised on strata fixed effects for each group separately and the group-specific mean is added to the residuals. The *p*-value for the Kolmogorov-Smirnov test for equality of distributions is reported. We also report the Randomization Inference *p*-value, obtained by permuting accreditation assignments for each facility 1,000 times.

Figure 13: Comparison of Accreditation Effects



Notes: The figure shows the effects of accreditation assignment, completion, and process on horizontal coordination (Panel (A)), vertical coordination (Panel (B)), expert weights meta-index (Panel (B)), and monthly mortality (Panel (C)). The accreditation process is measured by Harvard and NCCHC scores separately, which are detailed in Section VI.6. The coefficients of accreditation assignment come from estimating Equation 1. The coefficients of accreditation completion and process come from estimating Equation 2.

Figure 14: Heterogeneous Treatment Effects:
Reduced Form Estimates



Notes: This figure shows heterogeneity in treatment effects. This figure shows coefficients from the interaction between accreditation assignment and each heterogeneity factor predicted by experts from estimating Equation 1. Panel (A) and panel (B) are for the expert-weighted meta index and mortality, respectively.

Table 1: Quality Standard Indices

Category	Description	Examples
Standard A *Governance Administration*	Standards in this section address the establishment of a health care system that ensures access to care, professional administration of all aspects of health care, and monitoring and quality improvement policies that effectively process health care issues from identification through resolution.	• health record content • staff FTE • continuous quality improvement • grievance process
Standard B *Health Promotion, Safety and Disease Prevention*	Standards in this section address the need to optimize education, safety, and preventive care. Policies and procedures related to these standards require involvement by all facility staff.	• infectious disease prevention • suicide prevention • contraception
Standard C *Personnel Training*	Standards in this section ensure that appropriately trained personnel are in place to deliver health care to the inmate population and that qualified health care professionals are evaluated for continuing competency.	• staff credentials • custody staff training • health staff orientation
Standard D *Ancillary Health and Care Services*	Standards in this section address the establishment and maintenance of all necessary procedures for the provision of ancillary health care services.	• medications • clinic space and supplies • emergency response plan
Standard E *Patient Care and Treatment*	Standards in this section ensure the delivery of health care from arrival through discharge for health care issues. All care is timely, appropriate, and continues until resolution of the problem or until discharge.	• health care screenings • oral care • discharge planning • health care requests
Standard F *Special Needs and Services*	Standards in this section address patients with special health care needs and establish compliance requirements specific to each health care issue.	• mental health services • MAT • sexual abuse protocol • pregnancy care
Standard G *Medical-Legal Issues*	The standards in this section ensure that health services comply with legal requirements.	• use of restraints • segregation • informed consent/refusal

Table 2: Representativeness Table

	Mean	Coefficient	SE
Panel A: Jail Characteristics			
ADP	372.493	32.037	(54.743)
Admissions	5,177.080	1,281.462*	(747.622)
Avg. Stay Length (Months)	1.096	-0.207***	(0.079)
# Staff	104.897	33.541	(23.620)
Under Consent Decree	0.044	0.005	(0.034)
Maximum Jail Capacity	438.440	74.366	(86.527)
Holds Felony Offenders	0.484	-0.123	(0.077)
Panel B: Incarcerated Population Characteristics			
Share Male	0.833	0.011	(0.009)
Share White	0.596	-0.072**	(0.035)
Share Black	0.263	0.011	(0.035)
Share Hispanic	0.111	0.049	(0.032)
Share Felony	0.665	0.042	(0.031)
Share Misdemeanor	0.264	-0.013	(0.027)
F-stat		1.489	
Analytical Sample Facilities		41	
Comparative Facilities		1085	

Notes: Data are from the Census of Jails (COJ) (2019) (U.S. Department of Justice 2022). We compare jails in our analytical sample to other jails in the COJ. We match 41 out of the 44 jails who report in the COJ to 1085 other county jails that house adults with an ADP between 100 and 3000. Column (1) reports the sample mean. Column (2) reports the coefficient of an indicator, equal to one if the facility is within our sample. Column (3) reports the associated robust standard error. ADP, the number of admissions, and the number of staff are reported at the reporting unit level. A reporting unit can manage multiple facilities. We adjust such variables to the facility level by dividing by the number of facilities managed by each reporting unit. Average stay of length (months) is calculated as $(ADP * 365/Admissions)/12$. We report an omnibus test of balance by regressing in-sample status on all variables in the table, and compute the *F*-statistic from a test of the variables' joint significance. To preserve the full sample when computing the joint *F*-statistics, we replace missing values in each variable with zero and add a missing indicator. The *F*-statistic is reported at the bottom of the table. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: Balance Table

	Mean	Coefficient	SE
Panel A: Baseline Quality Standards			
Governance and Administration	0.502	-0.022	(0.041)
Health Promotion, Safety and Disease Prevention	0.493	0.069	(0.055)
Personnel and Training	0.504	0.000	(0.075)
Ancillary (Supportive) Health Care Services	0.502	0.006	(0.046)
Patient Care and Treatment	0.454	-0.009	(0.036)
Special Needs and Services	0.514	0.057	(0.062)
Medical-Legal Issues	0.544	0.017	(0.101)
Simple Mean Meta-Index	0.495	0.002	(0.040)
Principal Component Analysis Meta-Index	0.571	0.003	(0.067)
Expert Weights Meta-Index	0.495	0.016	(0.046)
F-Stat		1.865	
Panel B: Baseline Mortality			
Baseline Monthly Mortality in Death Logs	0.225	-0.401	(0.246)
Baseline Monthly Mortality in Media Reports	0.064	-0.035	(0.053)
Panel C: Baseline Staff Survey Responses			
# Staff Responses	25.864	-6.284	(7.461)
Share Male	0.517	0.038	(0.062)
Age	41.978	0.477	(1.511)
Share White	0.650	-0.023	(0.090)
Share Black	0.142	0.109	(0.075)
Share Hispanic	0.144	-0.047	(0.060)
Horizontal Coordination	0.656	-0.006	(0.030)
Vertical Coordination	0.671	-0.002	(0.040)
Satisfaction	0.687	-0.025	(0.021)
Respect Incarcerated Population	0.735	-0.005	(0.030)
F-Stat		1.345	
Panel D: Baseline Jail Characteristics			
ADP	384.326	-173.462	(122.271)
Avg. Stay Length (Months)	1.122	-0.001	(0.278)
Medical FTE	5.795	-1.180	(1.306)
Medical FTE Variety	5.818	-0.486	(0.962)
For-Profit Health Vendor	0.682	0.009	(0.144)
Previously Accredited	0.114	0.064	(0.095)
Share County Voting Republican	0.445	-0.021	(0.039)
Sheriff is Republican	0.318	0.018	(0.127)
F-Stat		1.320	
F-Stat		1.656	
Observations		44	
Strata FE		✓	

Notes: Data are from the baseline facility survey, baseline staff survey, medical audit look-back, death logs and media reports of death. Column (1) reports the sample mean, columns (2) and (3) report the coefficient on an indicator for treatment assignment and the associated robust standard error, respectively. Panel (A) compares baseline quality standards (for detailed definitions of the standards and their calculation see Appendix Section I). Panel (B) compares the baseline monthly mortality in death logs and in media reports. Share of health staff is the number of health staff who took the staff survey divided by the total number of staff that took the staff survey. ADP stands for average daily population. FTE is the number of full-time equivalents of health staff. For-profit health vendor indicates that the facility has hired an external private company to provide health services. Horizontal coordination, vertical coordination, satisfaction, and respect for inmates are staff sentiment indices constructed using individual questions from the staff survey. These questions are answered on a Likert scale and re-normalized on a scale from 0 to 1 with higher values representing stronger agreement. We report an omnibus test of balance by regressing treatment assignment on either the variables within the same panel or all the variables in the table, controlling for randomization strata fixed effects. "Baseline Monthly Mortality in Death Logs" is excluded from the omnibus test of balance to ensure a complete sample, as death logs are only available for 40 out of 44 facilities. We compute the *F*-statistic from a test of the variables' joint significance. Joint *F*-statistic is listed at the bottom of each panel and also at the bottom of the table. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Effect of Accreditation Completion on Staff Sentiment Indices: 2SLS Estimates

	Horizontal Coordination (1)	Vertical Coordination (2)	Satisfaction (3)	Respect Inmates (4)
	Panel A: Second Stage			
$\widehat{Accreditation}$	0.059**	0.037	0.037**	-0.010
	(0.027)	(0.031)	(0.018)	(0.023)
	Panel B: First Stage			
Treat	0.807***	0.782***	0.771***	0.784***
	(0.078)	(0.097)	(0.103)	(0.095)
F-stat	106.08	64.94	56.05	68.26
Control Complier Mean	0.618	0.627	0.639	0.764
Observations	671	671	671	671
Number of Clusters	44	44	44	44
Strata FE	✓	✓	✓	✓
Baseline Control	✓	✓	✓	✓

Notes: This table reports 2SLS estimates of the treatment effect of accreditation completion on four staff sentiment indices obtained from estimating Equation 2. All specifications include randomization strata fixed effects and the baseline mean of the respective outcome variable. Staff sentiment indices are created using data from the baseline/endline staff survey. The horizontal coordination index includes: (1) "Custody staff support the implementation of clinical decisions and partner with clinical staff"; (2) "Custody and clinical team staff work collaboratively to give good health care to inmates"; (3) "There is open communication between custody and health staff." The vertical coordination index includes: (1) "Staff feel comfortable speaking up when they see something that may negatively affect inmate care"; (2) "When staff speak up about inmate safety and care decisions, those with greater authority are open to their concerns"; (3) "My supervisor considers staff suggestions for improving inmate health and safety"; (4) "My supervisor considers staff suggestions for improving staff health and safety"; (5) "My supervisor takes direct actions to address staff concerns for improving inmate health and safety"; (6) "My supervisor takes direct actions to address staff concerns for improving staff health and safety." The satisfaction index includes: (1) "I received the training I need to do my job well"; (2) "I have the tools and resources I need to do my job well"; (3) "I am satisfied with my job"; (4) "I find my job to be meaningful"; (5) "I feel valued at my job"; (6) "I find my job to be challenging"; (7) "I would recommend my job to others." The respect inmates index includes: (1) "Inmates at this facility are treated with respect." These questions are answered on a Likert scale and renormalized on a scale from 0 to 1 with higher values representing stronger agreement. Control complier means are computed as described in the main text. Standard errors are clustered at the facility level. We obtained endline responses from 40 jails and we impute missing endline responses for the other 4 facilities in our sample with baseline values. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Table 5: Effect of Accreditation Completion on Medical Inputs:
2SLS Estimates

	Medical FTE (1)	Vendor Change (2)	Refer Out (3)	Equipment (4)	Services (5)
	Panel A: Second Stage				
$\widehat{Accreditation}$	-1.286	-0.063	-0.034	0.016	0.126**
	(0.949)	(0.203)	(0.036)	(0.032)	(0.051)
	Panel B: First Stage				
Treat	0.599***	0.578***	0.888***	0.580***	0.577***
	(0.101)	(0.104)	(0.070)	(0.102)	(0.105)
F-stat	35.27	31.04	161.58	32.29	30.32
Control Complier Mean	8.111	0.302	0.816	0.915	0.262
Observations	44	44	253	44	44
Baseline Control	✓		✓	✓	✓
Strata FE	✓	✓	✓	✓	✓

Notes: This table reports 2SLS estimates of the treatment effect of accreditation completion on medical inputs from estimating Equation 2. Column (1) is Medical FTE as reported in the facility survey. Column (2) is an indicator for whether the facility changed their vendor between baseline and endline as reported in the facility survey. Column (3) is the average of a rescaled Likert scale for whether health care staff can readily refer out. Column (4) covers medical equipment including: handwashing facilities/alternate means of hand sanitation, examination table, scale, thermometers, a light capable of providing direct illumination, blood pressure monitoring equipment, stethoscope, ophthalmoscope, otoscope, transportation equipment, trash containers for biohazardous materials and sharps, sterilizer for non-disposable medical or dental equipment, equipment and supplies for pelvic examinations if female inmates are housed in the facility, oxygen, AEDs, pulse oximeter, PPE, multiple-test dipstick urinalysis, finger-stick blook glucose tests, peak flow meters, stool blood-testing material and pregnancy test knits. In medical equipment, we also include dental equipment, hand washing facilities or alternate means of hand sanitation, dental examination chair, examination light for dental procedures, proper instruments for dental examination, trash containers for biohazardous materials and sharps from dental procedures, dentist's tools and PPE. Column (5) covers medical services including: pharmacy, records, radiology, physical therapy unit, emergency, labs, hospice and medical orthoses or prostheses or other aids to impairment. Each component of equipment and services is given equal weight, and the indices are calculated as the simple average share of all included items. All specifications except for Column (2) include the baseline measure and randomization strata fixed effects. Control complier means are computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Table 6: Effect of Accreditation Completion on Meta Quality Standards: 2SLS Estimates

	Simple Mean (1)	Principal Component (2)	Expert Weights (3)
	Panel A: Second Stage		
$\widehat{Accreditation}$	0.052	0.125**	0.085**
	(0.034)	(0.061)	(0.042)
	Panel B: First Stage		
Treat	0.575***	0.576***	0.564***
	(0.094)	(0.096)	(0.099)
F-stat	37.82	35.87	32.16
Control Complier Mean	0.599	0.632	0.572
Observations	44	44	44
Baseline Control	✓	✓	✓
Strata FE	✓	✓	✓

Notes: This table reports 2sls estimates of the treatment effect of accreditation completion on meta quality standards constructed by combining the seven quality standards in three ways as described in the main text: equal weights, the first component using principal component analysis, and reciprocal rank weights using our expert survey responses. For detailed definitions of the standards and their calculation see Appendix Section I and for expert weights see Figure 6. All specifications include the baseline quality standard and randomization strata fixed effects. Control complier means are computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Table 7: Effect of Accreditation Completion on Quality Standards:
2SLS Estimates

	Gov. Admin. (1)	Safety & Prevention (2)	Personnel Training (3)	Ancillary Services (4)	Patient Care & Rx. (5)	Special Needs (6)	Medical Legal (7)
	Panel A: Second Stage						
$\widehat{Accreditation}$	0.020	0.098*	0.143**	0.042	0.102**	0.021	0.114
	(0.031)	(0.059)	(0.071)	(0.031)	(0.047)	(0.078)	(0.078)
	Panel B: First Stage						
Treat	0.606***	0.533***	0.578***	0.575***	0.591***	0.544***	0.574***
	(0.088)	(0.106)	(0.104)	(0.104)	(0.092)	(0.100)	(0.103)
F-stat	47.40	25.21	31.12	30.63	41.01	29.52	31.16
Control Complier Mean	0.612	0.587	0.528	0.545	0.568	0.676	0.646
Observations	44	44	44	44	44	44	44
Baseline Control	✓	✓	✓	✓	✓	✓	✓
Strata FE	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports 2sls estimates of the treatment effect of accreditation completion on the seven quality standards from estimating Equation 2 (for detailed definitions of the standards and their calculation see Appendix Section I). All specifications include the baseline quality standard and randomization strata fixed effects. Control complier means are computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Table 8: Effect of Accreditation Completion on Mortality: 2SLS Estimates

	(1)	(2)	(3)
	Panel A: Second Stage		
$\widehat{Accreditation}$	-0.296***	-0.225***	-0.249***
	(0.114)	(0.073)	(0.092)
	Panel B: First Stage		
Treat	0.578***	0.603***	0.568***
	(0.104)	(0.103)	(0.108)
F-stat	31.04	34.12	27.71
Control Complier Mean	0.333	0.262	0.288
Observations	44	44	44
Strata FE	✓	✓	✓
Control for ADP		✓	
Baseline Control			✓

Notes: This table reports 2SLS estimates of the effect of accreditation completion on mortality obtained from estimating Equation 2. The outcome variable is the average number of deaths per month occurring in the six-month window starting ten months after treatment assignment. Column (1) controls for randomization strata fixed effects. Column (2) adds a control for the log ADP of the jail. Column (3) replaces log ADP with baseline mortality. Baseline mortality refers to the average number of deaths per month occurring during the six months preceding treatment assignment. All specifications include randomization strata fixed effects. Control complier means are computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Table 9: Effect of Accreditation Completion on Recidivism and Litigation:
2SLS Estimates

	Recidivism			Litigation		
	Has Jail Records	3-Month Recid	6-Month Recid	# Lawsuits		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Second Stage					
$\widehat{Accreditation}$	-0.016	-0.128	-0.208**	-0.222	-0.223	-0.222
	(0.216)	(0.084)	(0.086)	(0.157)	(0.141)	(0.154)
	Panel B. First Stage					
Treat	0.578***	0.662***	0.621***	0.578***	0.603***	0.582***
	(0.104)	(0.131)	(0.130)	(0.104)	(0.103)	(0.104)
F-stat	31.04	25.65	22.88	31.04	34.12	31.46
Control Complier Mean	0.778	0.238	0.397	0.302	0.286	0.299
Observations	44	9748	9748	44	44	44
Strata FE	✓	✓	✓	✓	✓	✓
Control for ADP					✓	
Baseline Control						✓

Notes: This table reports 2SLS estimates of the effect of accreditation completion recidivism and litigation. Column 1 shows the treatment effects of accreditation completion on whether having booking records. Columns 2 and 3 show the treatment effects on 3-month and 6-month recidivism, with the outcome variable being a binary indicator for whether an inmate released within 45 days of the endline survey is re-booked within 3 months and 6 months of release, respectively. Columns 2-3 control for the baseline 3-month or 6-month recidivism rate, respectively. Column 4-6 show the treatment effects of accreditation completion on healthcare related lawsuits. The outcome variable is the total number of lawsuits occurring in the six-month window following the endline survey. Column 5 controls for log ADP. Column 6 controls for the baseline lawsuits instead, which refers to the total number of lawsuits occurring during the six months preceding the baseline survey. All specifications include randomization strata fixed effects. Control complier means are computed as described in the main text. Columns 2-3's standard errors are clustered at the facility level. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix

Table of Contents

A Tables and Figures **A.3**

B Model of Accreditation **A.29**

B.1 No Accreditation . . . A.30

B.2 Accreditation as Coordination . . . A.30

C Recruitment **A.32**

D Enrollment **A.33**

E Randomization and Stratification **A.34**

F Accreditation Process Initiation and Completion **A.35**

G Description of Data Elements **A.36**

G.1 Facility Survey . . . A.36

G.2 Staff Survey . . . A.36

G.3 Medical Audit . . . A.36

G.4 Leadership Interviews . . . A.37

G.5 Inmate Interviews . . . A.37

G.6 Expert Prediction Survey . . . A.37

H Coding **A.40**

H.1 Coding Committees and Codebook . . . A.40

H.2 Variable Transformation . . . A.40

H.3 Coding Outcomes . . . A.42

H.4 Coding Facility Survey . . . A.42

H.5 Coding Staff Survey . . . A.42

H.6 Coding Medical Audits . . . A.42

H.7 Coding Litigation . . . A.43

I Construction of Indices **A.44**

J Cost-Benefit Analysis **A.45**

J.1 Benefits of Accreditation . . . A.45

J.2 Costs of Accreditation . . . A.45

J.3 Cost-Benefit Calculation . . . A.45

K Deviations from Pre-Analysis Plan **A.47**

L Institutional Appendix **A.48**

L.1 Jails A.48
L.2 Quality Standards in Jails A.48
L.3 NCCHC A.49
L.4 Legislation A.50
L.5 Evidence on Accreditation and Health A.50

M Additional Materials **A.52**
M.1 Index of Items A.52

A Tables and Figures

Appendix Figure A1: Recruitment Email

Subject Line: Healthcare Accreditation Opportunity

Dear _____,

Congratulations! Your facility has been selected as an excellent candidate to participate in a study at Harvard University. The National Commission of Correctional Healthcare (NCCHC) has partnered with Harvard University to study the effects of accreditation on health care delivery in jails, and we are currently recruiting facilities **that are not currently accredited with the NCCHC to undergo healthcare accreditation.**

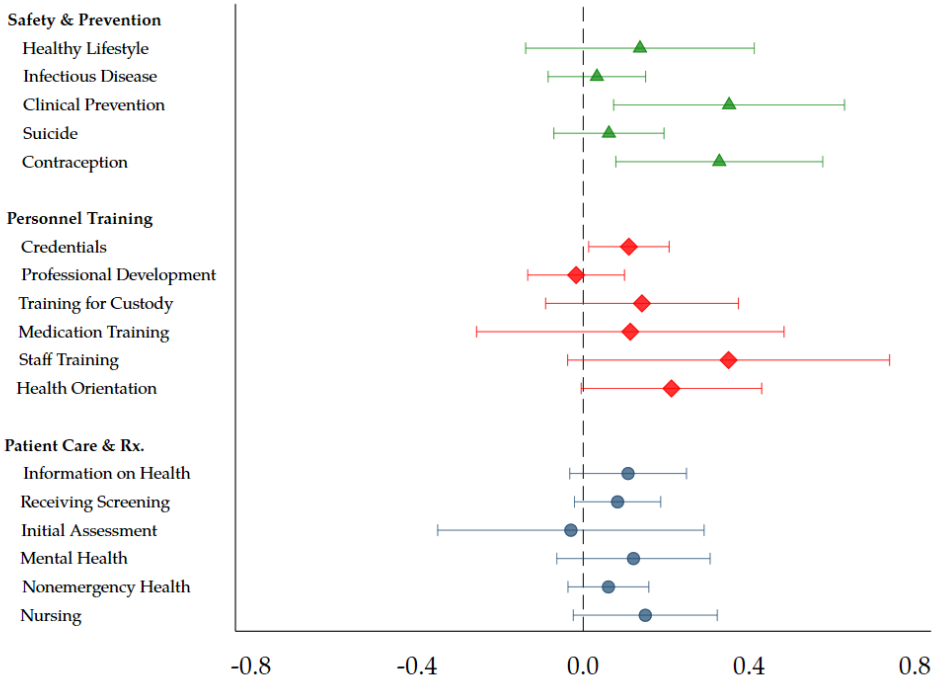
This is a terrific opportunity to have the health care provided in your facility accredited by the National Commission on Correctional Health Care (NCCHC) with a substantial discount. Participation involves completing surveys at two timepoints for which facilities will be compensated up to $1,000, an *optional* confidential assessment of the health care delivery system from Harvard University researchers, and healthcare accreditation by the NCCHC at an expedited timeline and substantially discounted rate.

Our study has been supported by the American Jail Association, approved by the Harvard Research Ethics Board, and granted a Certificate of Confidentiality from the National Institutes of Health. We attach: (1) a flyer with more study information; (2) our letter of support from the American Jail Association; (3) our institutional approval from the Harvard research ethics board; as well as (4) our Certificate of Confidentiality from the National Institute of Health, which guarantees complete protection of privacy of all participating facilities in our study, including during civil/criminal/administrative/legislative proceedings. Please feel free to forward these materials to any interested parties.

Would a ½-hour informational meeting to discuss this study work for you next week? Our schedules are extremely flexible, and we are currently recruiting sites, so contact us at your earliest convenience at accredstudy@hks.harvard.edu or by calling [NAME] at [PHONE].

Thank you for your commitment to correctional healthcare.

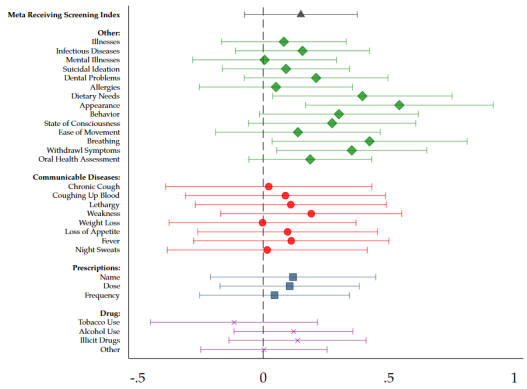
Appendix Figure A2: Effect of Accreditation Completion on Sub-Standards
2SLS Estimates



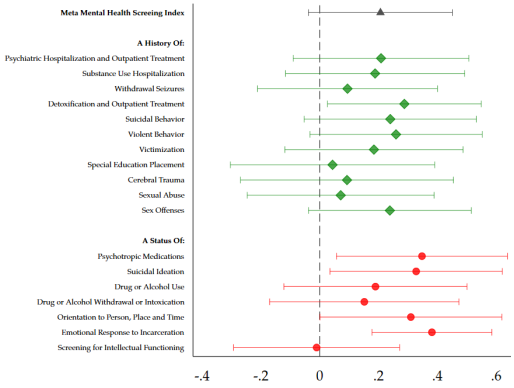
Notes: This figure shows the 2SLS estimates of the effect of the accreditation completion on each subsection under quality standard B, C and E from estimating Equation 2 (for detailed definitions of the standards and their calculation see Appendix Section I). Under Safety and Prevention Standard, "Healthy Lifestyle" subsection refers to "Health care policies, procedures, and practices emphasize health promotion, wellness, and recovery". "Infectious Disease" subsection refers to "There is a comprehensive institutional program that includes surveillance, prevention, and control of communicable disease". "Clinical Prevention" subsection refers to "Inmates are provided with clinical preventive services as medically indicated". "Suicide" subsection refers to "Suicides are prevented when possible by implementing prevention efforts and intervention". "Contraception" subsection refers to "Contraception is made available as clinically indicated". Under Personnel Training Standard, "Credentials" subsection refers to "The facility's qualified health care professionals are legally eligible to perform their clinical duties". "Professional Development" refers to "The facility's qualified health care professionals maintain current clinical knowledge and skills". "Training in Custody" subsection refers to "Correctional officers are trained to recognize the need to refer an inmate to a qualified health care professional". "Medication Training" subsection refers to "Personnel who administer or deliver prescription medication are appropriately trained". "Staff Training" subsection refers to "Does the provided documentation show that there is a documented plan for custody staff to follow when a health situation arises and health staff are not present". "Health Orientation" subsection refers to "Health staff are properly acclimated to work in the correctional environment and understand their roles and responsibilities". Under Patient Care and Rx. Standard, "Information on Health" refers to "Upon arrival at the facility, inmates are informed of the availability of health care services and how to access them". "Receiving Screening" subsection refers to "Screening is performed on all inmates upon arrival at the intake facility to ensure that emergent and urgent health needs are met". "Initial Assessment" subsection refers to "Inmates receive initial health assessments". "Mental Health" subsection refers to "Mental health screening is performed to ensure that urgent mental health needs are met". "Nonemergency Health" refers to "Inmates' nonemergent health care needs are met". "Nursing" subsection refers to "Nursing assessment protocols and procedures are appropriate to the level of competency and preparation of the nursing personnel and comply with the relevant state practice acts". 95% confidence interval is plotted.

Appendix Figure A3: Effect of Accreditation Completion on Screening
2SLS Estimates

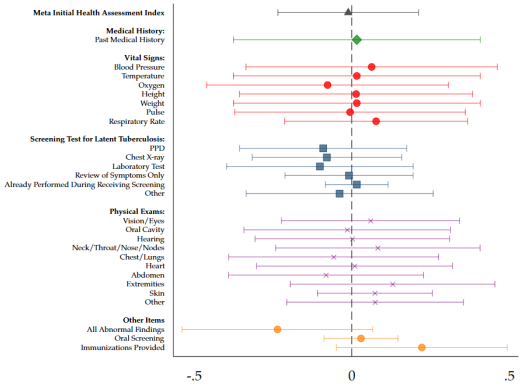
Panel (A) Receiving Screening



Panel (B) Mental Health Screening

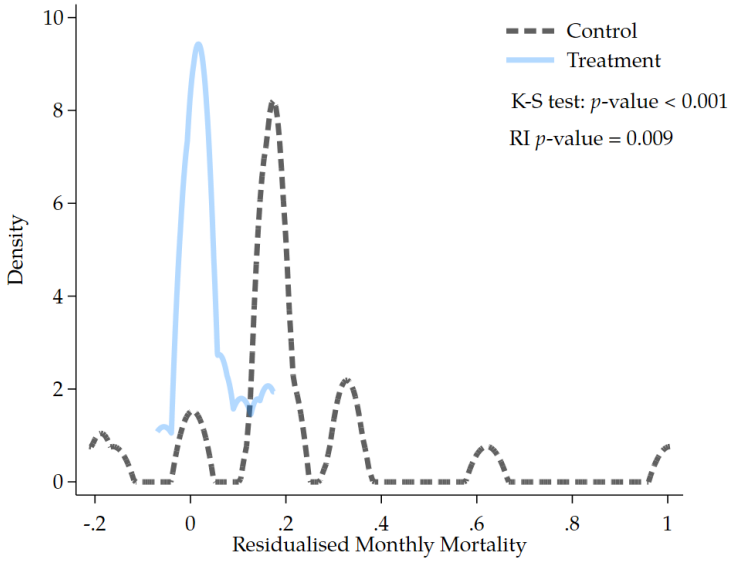


Panel (C) Initial Health Assessment



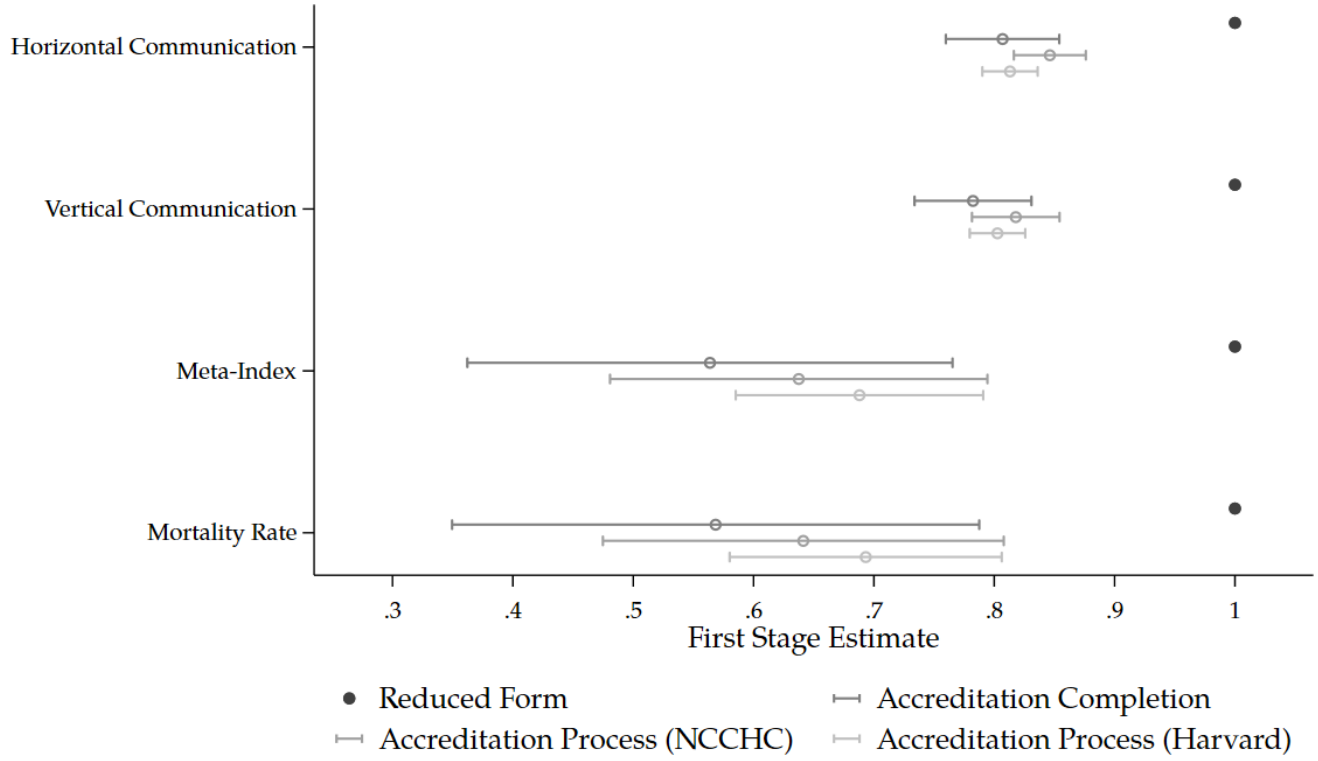
Notes: The figure shows the 2SLS estimates of the effect of the accreditation completion on each component under subsections "Receiving Screening" (Panel (A)), "Mental Health Screening" (Panel (B)), and "Initial Health Assessment" (Panel (C)). 95% confidence interval is plotted. In each figure, the black triangle represents the meta index by taking the average of all the components within the subsection.

Appendix Figure A4: Effect of Accreditation Assignment on Mortality: Starting Sample



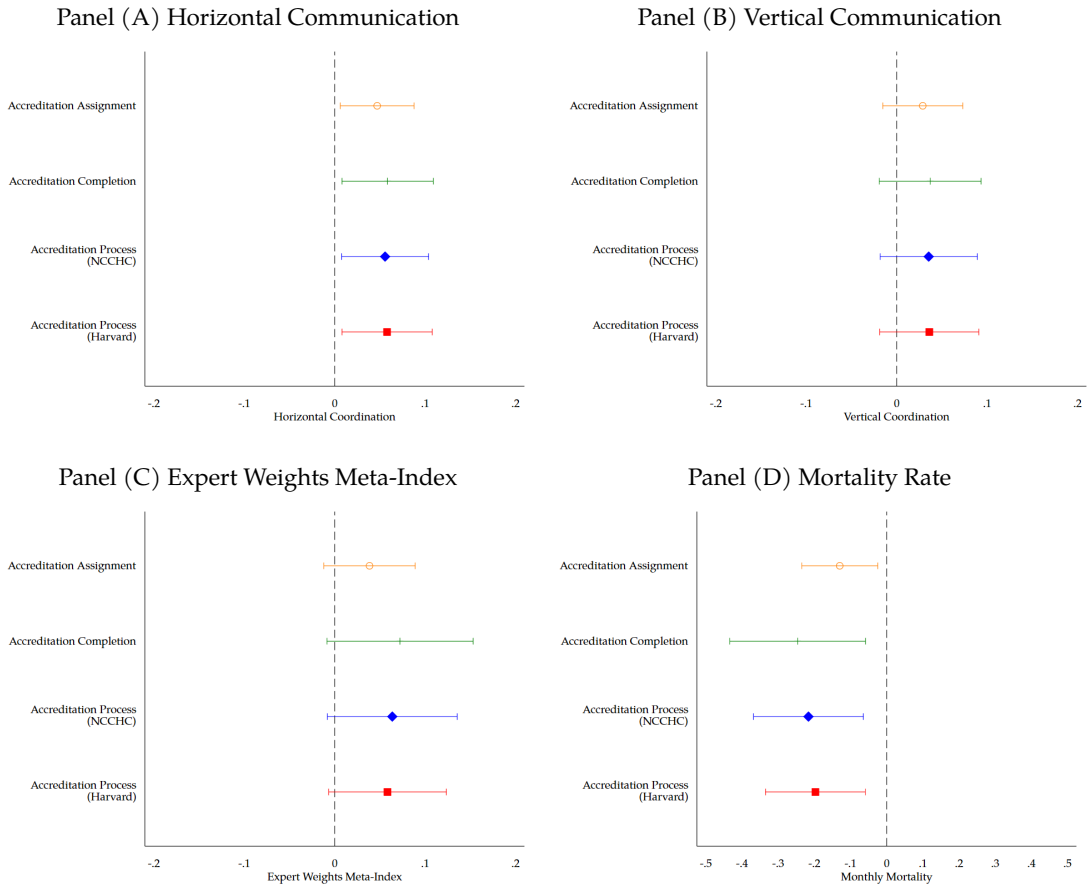
Notes: This figure shows the distribution of mortality before and after accreditation between treatment and control facilities. Mean monthly mortality is residualised on strata fixed effects for each group separately and the group-specific mean is added to the residuals. The *p*-value for the Kolmogorov-Smirnov test for equality of distributions is reported. We also report the Randomization Inference *p*-value, obtained by permuting accreditation assignments for each facility 1,000 times.

Appendix Figure A5: Comparison of First Stage Estimates for Accreditation



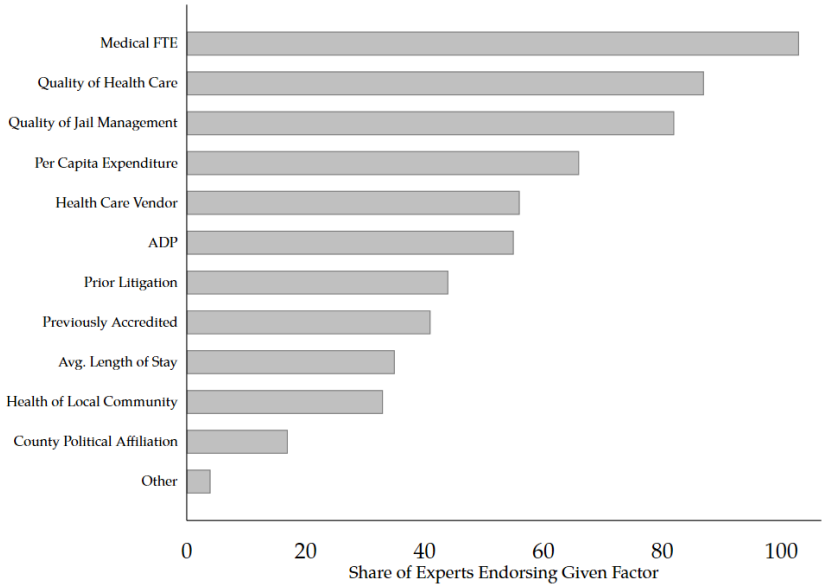
Notes: The figure shows the first stage estimates of accreditation assignment on accreditation completion and process. The accreditation process is measured by Harvard and NCCHC scores separately, which are detailed in Section VI.6. The coefficients come for estimating the first stage of Equation 2. For the reduced form, an implied coefficient of 1 is plotted for reference.

Appendix Figure A6: Comparison of Accreditation Effects: Starting Sample



Notes: The figure shows the effects of accreditation assignment, completion and process on horizontal communication (Panel (A)), expert weights meta-index (Panel (B)) and monthly mortality (Panel (C)). The accreditation process is measured by Harvard and NCCHC scores separately, as detailed in Section VI.6. The coefficients of accreditation assignment come from estimating Equation 1. The coefficients of accreditation completion and process come from estimating Equation 2.

Appendix Figure A7: Expert Predictions and Heterogeneous Treatment Effects



Notes: Respondents are asked "Which of the following factors do you think will impact whether NCCHC accreditation of jail health care services will affect the outcomes of interest? Recall, our outcomes of interest include health care procedures, health outcomes, staff outcomes, and litigation against jails." The figure plots the share of all respondents choosing a given factor; they could select up to five factors listed in the graph.

Appendix Figure A8: Word Cloud of N-Grams From Jail Leadership Interviews Regarding Experience with Accreditation

Panel (A) Experience Undergoing Accreditation



Panel (B) Challenges Faced During Accreditation



Panel (C) Biggest Change Implemented



Panel (D) Relationship between Custody and Medical Staff



Panel (E) Changes in Health Care Delivery



Notes: This figure shows word clouds of n-grams from interviews with jail leadership of accredited facilities who were asked 5 questions about their experience undergoing accreditation. The specific questions/stems included: Panel (A) ''Please describe your facility's experience undergoing NCCHC accreditation process." ; Panel (B) ''Were there any challenges your facility faced during the NCCHC accreditation process? If so, what were they?" ; Panel (C) ''In your opinion, what is the biggest change your facility has implemented as a result of the NCCHC accreditation process?"; and Panel (D) ''Do you think that the relationship and coordination between correctional and health care staff has changed due to the NCCHC accreditation process? If so, how has it changed?" Each word cloud is generated after removing commonly occurring filler words in these quotes. Panel (B) additionally filters out all n-grams with the word "challenge" as it is part of the question and quotes usually begin with "the challenges was..." or a similar phrase; Panel (D) additionally filters out all n-grams with the word "relationship" as it is part of the question and quotes usually begin with "the relationship..." or a similar phrase. The word "anything" is also filtered; and Panel (E) additionally filters out all n-grams with the word "health care" as it is part of the question.

Appendix Table A1: Essential and Important Subsections

Category	Subsections
Essential Standards	A01 - Access to Care
	A02* - Responsible Health Authority
	A03 - Medical Autonomy
	A04* - Administrative Meetings and Reports
	A05* - Policies and Procedures
	A06* - Continuous Quality Improvement Program
	A08* - Health Records
	B02* - Infectious Disease Prevention and Control
	B03* - Clinical Preventive Services
	B05* - Suicide Prevention and Intervention
	B07 - Communication on Patients' Health Needs
	C01* - Credentials
	C03* - Professional Development
	C04* - Health Training for Correctional Officers
	C05* - Medication Administration Training
	C06 - Inmate Worker
	C08 - Health care Liaison
	D01 - Pharmaceutical Operations
	D02* - Medication Services
	D05* - Medical Diets
	D07* - Emergency Services and Response Plan
	D08 - Hospital and Specialty Care
	E01* - Information on Health Services
	E02* - Receiving Screening
	E04* - Initial Health Assessment
	E05* - Mental Health Screening and Evaluation
	E07* - Nonemergency health care Requests and Services
	E09 - Continuity, Coordination, and Quality of Care During Incarceration
	E10 - Discharge Planning
	F01* - Patients with Chronic Disease and Other Special Needs
	F02 - Infirmary-Level Care
	F03* - Mental Health Services
	F04* - Medically Supervised Withdrawal and Treatment
	F05* - Counseling and Care of the Pregnant Inmate
	F06* - Response to Sexual Abuse
	G01* - Restraint and Seclusion
	G02* - Segregated Inmates
	G03* - Emergency Psychotopic Medication
Important Standards	A07* - Privacy of Care
	A09* - Procedure in the Event of an Inmate Death
	A10* - Grievance Process for health care Complaints
	B01* - Healthy Lifestyle Promotion
	B04 - Medical Surveillance of Inmate Workers
	B06* - Contraception
	B08 - Patient Safety
	B09 - Staff Safety
	C02 - Clinical Performance Enhancement
	C07* - Staffing
	C09* - Orientation for Health Staff
	D03* - Clinic Space, Equipment, and Supplies
	D04* - On-Site Diagnostic Services
	D06 - Patient Escort
	E03 - Transfer Screening
	E06 - Oral Care
	E08* - Nursing Assessment Protocols and Procedures
	F07* - Care for the Terminally Ill
	G04* - Therapeutic Relationship, Forensic Information, and Disciplinary Actions
	G05 - Informed Consent and Right to Refuse
	G06 - Medical and Other Research

Notes: This table shows the NCCHC classifications of subsections into essential or important standards. The asterisk indicates whether the subsection is measured using the Harvard study metrics.

Appendix Table A2: Subsidies for Accreditation by Treatment Status

	Treatment	Control
Subsidy for Accreditation	$7,110.83	$2,181.82
Survey Incentives	$813.64	$1,104.55
Total Amount	$7,924.47	$3,286.36

Notes: "Subsidy for Accreditation" includes the total contribution towards accreditation. For both groups, "Survey Incentives" includes incentives for finishing facility survey and staff survey. For facility surveys, facilities can opt-in to receive $500 in direct payment via check or direct deposit, or can credit the payment towards their accreditation or re-accreditation at endline. For staff surveys, staff who complete the survey are entered into a lottery for a $100 gift card. Additional details can be found in Appendix Section H.5. "Total Amount" is the sum of "Subsidy for Accreditation" and "Survey Incentives". The difference between Treatment and Control for total amount is statistically significant (p-value<0.001).

Appendix Table A3: Effect of Accreditation Assignment on Initiation and Completion Rate

		Endline Study Components				
	Failure to Initiate (1)	Average Completed (2)	Facility Survey (3)	Medical Audits (4)	Death Logs (5)	Staff Survey (6)
	Panel A: 2SLS Estimates					
$\widehat{Accreditation}$	0.124	0.098	0.008	0.194	0.269*	-0.079
	(0.089)	(0.122)	(0.158)	(0.205)	(0.163)	(0.160)
	Panel B: Reduced Form Estimates					
Treat	0.067	0.053	0.004	0.104	0.144	-0.042
	(0.047)	(0.073)	(0.092)	(0.121)	(0.093)	(0.090)
Control Mean	0.00	0.83	0.86	0.73	0.82	0.91
Observations	46	46	46	46	46	46
Strata FE	✓	✓	✓	✓	✓	✓

Notes: This table shows the treatment effects of accreditation assignment on failure to initiate the accreditation process and completion rate of either a specific endline survey (Columns (3) – (6)) or the average completion of endline components in Column (2). Panel (A) shows 2SLS estimates from Equation 2, and Panel (B) shows reduced form estimates from Equation 1. Robust standard errors are reported in parentheses. Asterisks indicate coefficient statistical significance level : $*p < 0.10; **p < 0.05; ***p < 0.01$.

Appendix Table A4: Balance Table on Enrollment

	Overall Mean	Difference	SE
	(1)	(2)	(3)
Monthly Mortality at Screening	0.057	0.001	(0.035)
2019 ADP	645.793	-396.294	(252.836)
Has Health Vendor	0.786	0.152	(0.151)
NaphCare/Wellpath	0.179	0.121	(0.102)
For-Profit Health Vendor	0.875	0.159	(0.135)
Sheriff Up For Election	0.482	-0.023	(0.166)
Anticipated Vendor Change	0.146	-0.214	(0.203)
Enrolled Facilities		44	
Excluded Facilities		12	

Notes: Column (1) reports the sample mean. Column (2) reports the coefficient of regressing an indicator, equal to one if the facility enrolled in the experiment. Column (3) reports the associated robust standard error. Monthly mortality at screening refers to the average number of deaths in media reports in the 6-month period ranging from 2 months prior to screening up to 4 months post-screening. For-profit health vendor is defined as whether the health vendor is a private company. $*p < 0.10; **p < 0.05; ***p < 0.01$.

Appendix Table A5: Effects of Accreditation on Components of Staff Sentiment Indices: Reduced Form and 2SLS Estimates

	Reduced Form (1)	2SLS (2)
Panel A: Horizontal Coordination Index		
Custody Support Health's Decisions	0.053**	0.067**
	(0.024)	(0.032)
Custody and Health Coordinate	0.039*	0.048*
	(0.022)	(0.026)
Custody and Health Communicate	0.051**	0.064**
	(0.024)	(0.031)
Panel B: Vertical Coordination Index		
Staff Comfortable Speaking Up on Inmate Care	0.067**	0.075**
	(0.034)	(0.037)
Authority Open to Concerns on Inmate Care	0.089***	0.101***
	(0.028)	(0.032)
Supr. Open to Concerns on Inmate	0.011	0.014
	(0.029)	(0.037)
Supr. Open to Concerns on Staff	0.028	0.035
	(0.027)	(0.034)
Supr. Consider Suggestions on Inmate	0.028	0.035
	(0.032)	(0.040)
Supr. Consider Suggestions on Staff	0.049*	0.063**
	(0.025)	(0.031)
Panel C: Satisfaction Index		
Properly Trained	0.022	0.029
	(0.022)	(0.029)
Sufficient Tools	0.043*	0.056*
	(0.023)	(0.030)
Satisfied	0.030	0.039
	(0.021)	(0.026)
Meaningful	0.012	0.015
	(0.022)	(0.028)
Valued	0.053*	0.069*
	(0.029)	(0.037)
Chanllenging	0.004	0.005
	(0.021)	(0.026)
Recommend to Others	0.038*	0.050*
	(0.023)	(0.029)

Notes. This table shows the treatment effect of accreditation on the components of staff sentiment indices. Column (1) reports reduced form estimates from Equation 1. Column (2) reports 2SLS estimates from Equation 2, using the accreditation completion. All specifications include randomization strata fixed effects and the baseline mean of the respect outcome variable at the facility level. We obtained endline responses from 40 jails and we impute missing endline responses for the other 4 facilities in our sample with baseline values. Standard errors are clustered at the facility level. Asterisks indicate coefficient statistical significance level: $*p < 0.10; **p < 0.05; ***p < 0.01$.

Appendix Table A6: Effect of Accreditation Completion on Meta-Index by Expert Type: 2SLS Estimates

			Expert Weights	
	Simple Mean	Principal Component	All Other Experts	Formerly Incarcerated
	(1)	(2)	(3)	(4)
	Panel A: Second Stage			
$\widehat{Accreditation}$	0.052	0.125**	0.084**	0.087**
	(0.034)	(0.061)	(0.042)	(0.042)
	Panel B: First Stage			
Treat	0.575***	0.576***	0.563***	0.566***
	(0.094)	(0.096)	(0.099)	(0.100)
F-stat	37.82	35.87	32.11	32.29
Control Complier Mean	0.599	0.632	0.572	0.571
Observations	44	44	44	44
Baseline Control	✓	✓	✓	✓
Strata FE	✓	✓	✓	✓

Notes. This table reports 2SLS estimates of the effect of accreditation completion on the meta-index of quality standards obtained from estimating Equation 2. Detailed definitions of the standards and expert weights can be found in Appendix Section I and Figure 6, respectively. All specifications include the baseline quality standard and randomization strata fixed effects. The outcome variable is the endline meta index. Control complier mean is computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A7: Effect of Accreditation Process on Consolidated Standards: 2SLS Estimates

	Essential Standard	Important Standard
	(1)	(2)
	Panel A: Second Stage	
$\widehat{Accreditation}$	0.043	0.130**
	(0.031)	(0.065)
	Panel B: First Stage	
Treat	0.579***	0.563***
	(0.094)	(0.097)
F-stat	29.99	32.01
Control Complier Mean	0.603	0.565
Observations	44	44
Baseline Control	✓	✓
Strata FE	✓	✓

Notes: This table reports 2SLS estimates of the treatment effect of accreditation process on consolidated essential and important standards obtained from estimating Equation 2. All specifications include the baseline consolidated standards and randomization strata fixed effects. The outcome variable is the endline consolidated quality, constructed based on the classification shown in Appendix Table A5. Control complier mean is computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A8: Effect of Accreditation Completion on Morbidity Screening and Treatment: 2SLS Estimates

	Screening Index			Treatment Index		
	SI	OUD	HTN	SI	OUD	HTN
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Second Stage					
$\widehat{Accreditation}$	0.242**	0.207*	0.011	0.087	-0.074	0.086
	(0.114)	(0.115)	(0.193)	(0.105)	(0.104)	(0.120)
	Panel B: First Stage					
Treat	0.587***	0.575***	0.556***	0.562***	0.566***	0.496***
	(0.096)	(0.097)	(0.104)	(0.104)	(0.100)	(0.107)
F-stat	37.03	34.99	28.80	29.42	32.35	21.65
Control Complier Mean	0.513	0.451	0.561	0.757	0.820	0.480
Observations	88	88	88	88	88	88
Number of Clusters	44	44	44	44	44	44
Strata FE	✓	✓	✓	✓	✓	✓
Baseline Control	✓	✓	✓	✓	✓	✓

Notes: This table reports 2SLS estimates of the treatment effect of accreditation completion on medical education from estimating Equation 2. Columns (1)-(3)'s outcome variable is a screening index specific for each illness. We match the items checked in the screening with each type of illness. SI screening index includes "history of or current suicidal ideation," "a history of suicidal behavior," and "the status of suicidal ideation." OUD screening index includes "current or prior withdrawal symptoms," "illicit drugs," "a history of substance use hospitalization," "a history of withdrawal seizures," "a history of detoxification and outpatient treatment," "the status of drug or alcohol use," and "the status of drug of alcohol withdrawal intoxication." HTN screening index includes "blood pressure," "pulse," and "respiratory rate." Columns (4)-(6)'s outcome variable is a treatment index specific for each illness. SI treatment index includes "Do health records document that they received education (related to suicidal ideation)?" "For patients with suicidal ideation, is there evidence of a treatment plan?" "For patients with suicidal ideation, is there evidence that the inmates were monitored for suicide risk?" and "Whether custody staff or health staff are monitoring patients with suicidal ideation?" OUD treatment index includes "Do health records document that they received education on MAT?" "Are protocols used for the management of withdrawal?" and "Is follow-up plan and treatment documented?" HTN treatment index includes "Do health records document that they received education on the following? Salt intake/Weight loss/Medication use and frequency" "Do orders include low sodium diet/duration/specific instructions?" "Does blood pressure appear to be controlled based on most recent reading?" and "Is follow-up plan and treatment documented?" All specifications include the baseline measure and randomization strata fixed effects. Control complier means are computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A9: Effect of Accreditation Completion on Probability of Death within U.S. Jails: 2SLS Estimates

	Probability that Number of Deaths Equals		
	[0-1]	[2-4]	$[\geq 5)$
	(1)	(2)	(3)
$\widehat{Accreditation}$	0.481***	-0.323***	-0.100
	(0.143)	(0.120)	(0.085)
Control Complier Mean	0.524	0.302	0.175
F-stat	30.63	30.46	27.83
Observations	44	44	44
Strata FE	✓	✓	✓
Baseline Control	✓	✓	✓

Notes: This table reports 2SLS estimates of the treatment effect of accreditation completion on mortality obtained from estimating Equation 2. All specification include randomization strata fixed effects. The outcome variable is an indicator that takes the value of 1 if the interval it represents includes the total number of deaths occurring in the six-month window starting ten months after treatment assignment. Similar indicators are created for the total number of deaths that take place in the 6 months before treatment assignment and included as a control. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A10: Effect of Accreditation Completion on Mortality:
2SLS Estimates

	Count Model						
	Poisson		Negative-Binomial		Log Mortality		Mortality/ ADP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\widehat{Accreditation}$	-2.411***	-2.375***	-2.243**	-2.181**	-0.555***	-0.557***	-0.521**
	(0.786)	(0.872)	(0.871)	(0.942)	(0.184)	(0.179)	(0.208)
Control Mean	1.227	1.227	1.227	1.227	0.552	0.552	0.626
Observations	44	44	44	44	44	44	44
Strata FE	✓	✓	✓	✓	✓	✓	✓
Control for ADP	✓	✓	✓	✓	✓	✓	
Baseline Control		✓		✓		✓	✓

Notes. This table shows the 2SLS estimates of the treatment effect of accreditation completion on mortality, where accreditation is instrumented by accreditation assignment. Columns (1)-(2) report estimates from a Generalized Linear Model (GLM) with Poisson distributed errors and a log-link function. Columns (3)-(4) report GLM estimates with negative-binomial distributed errors and a log-link function. Columns (1)-(4) use the log of current Average Daily Population (ADP) as the offset. Columns (5)-(6) use the log mortality (1 is added to mortality prior to the log transformation) and column (7) uses the mortality rate (per 1000 inmates). All specifications include strata fixed effect (since randomization occurred within a combination of batch and ADP size group). The sample includes 44 jails recruited into the study. Baseline mortality refers to the total number of deaths occurring during the six months preceding treatment assignment. The outcome variable is the total number of deaths occurring in the six-month window starting ten months after treatment assignment. In columns (2), (4), and (6) we control for baseline level of mortality. Standard errors are robust. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A11: Effect of Accreditation on Mortality by Source of Data: Reduced Form Estimates

	Death Logs	Media Reports	Union of Logs and Media
	(1)	(2)	(3)
Treat	-0.228**	-0.081***	-0.141***
	(0.091)	(0.031)	(0.053)
Control Mean	0.233	0.121	0.205
Observations	37	44	44
Strata FE	✓	✓	✓
Baseline Control		✓	✓

Notes. This table reports reduced form estimates of the treatment effect of accreditation assignment on mortality obtained from estimating Equation 1. All specifications include randomization strata fixed effects and baseline mean. Baseline mortality refers to the average number of deaths per month occurring during the six months preceding treatment assignment. The outcome variable is the average number of deaths per month occurring in the six-month window starting ten months after treatment assignment. Column (1) uses the union of the total number of deaths in the death logs and media reports as the outcome. Column (2) uses only the total number of deaths as per media reports as the outcome. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A12: Effect of Accreditation Completion on Mortality by Cause: 2SLS Estimates

	Illness	Suicide & Homicide	Undetermined	Drug Overdose
	(1)	(2)	(3)	(4)
	Panel A: Second Stage			
$\widehat{Accreditation}$	-0.984**	-0.603	-1.492	-0.286
	(0.400)	(0.465)	(1.108)	(0.213)
	Panel B: First Stage			
Treat	0.578	0.578	0.578	0.578
	0.106	0.106	0.106	0.106
F-stat	29.99	29.99	29.99	29.99
Control Complier Mean	1.063	0.905	1.635	0.508
Observations	44	44	44	44
Strata FE	✓	✓	✓	✓

Notes: This table reports 2SLS estimates of the effect of accreditation completion on mortality obtained from estimating Equation 2, by cause of death. The outcome variable is the total number of deaths of each cause reported in the 12 months at endline. All specifications include randomization strata fixed effects. Control complier means are computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A13: Effect of Accreditation Assignment on Recidivism and Litigation: Reduced Form Estimates

	Recidivism			Litigation		
	Has Jail Records	3-Month Recid	6-Month Recid	# Lawsuits		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.009	-0.085	-0.129**	-0.128	-0.135	-0.129
	(0.136)	(0.056)	(0.058)	(0.094)	(0.090)	(0.094)
Control Mean	0.727	0.194	0.289	0.182	0.182	0.182
Observations	44	9748	9748	44	44	44
Strata FE	✓	✓	✓	✓	✓	✓
Control for ADP					✓	
Baseline Control						✓

Notes: This table reports RF estimates of the effect of accreditation assignment on recidivism and litigation. Column 1 shows the treatment effects of accreditation assignment on whether having booking records. Columns 2 and 3 show the treatment effects on 3-month and 6-month recidivism, with the outcome variable being a binary indicator for whether an inmate released within 45 days of the endline survey is re-booked within 3 months and 6 months of release, respectively. Columns 2-3 control for the baseline 3-month or 6-month recidivism rate, respectively. Column 4-6 show the treatment effects of accreditation assignment on healthcare related lawsuits. The outcome variable is the total number of lawsuits occurring in the six-month window following the endline survey. Column 5 controls for log ADP. Column 6 controls for the baseline lawsuits instead, which refers to the total number of lawsuits occurring during the six months preceding the baseline survey. All specifications include randomization strata fixed effects. Columns 2-3's standard errors are clustered at the facility level. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A14: Balance Table, Starting Sample

	Mean	Coefficient	SE
Panel A: Baseline Quality Standards			
Governance and Administration	0.494	-0.028	(0.040)
Health Promotion, Safety and Disease Prevention	0.480	0.049	(0.056)
Personnel and Training	0.491	-0.011	(0.072)
Ancillary (Supportive) Health Care Services	0.490	-0.009	(0.046)
Patient Care and Treatment	0.446	-0.013	(0.035)
Special Needs and Services	0.500	0.034	(0.061)
Medical-Legal Issues	0.527	-0.007	(0.102)
Simple Mean Meta-Index	0.485	-0.008	(0.039)
Principal Component Analysis Meta-Index	0.557	-0.012	(0.066)
Expert Weights Meta-Index	0.483	0.003	(0.046)
F-Stat		1.075	
Panel B: Baseline Mortality			
Baseline Monthly Mortality in Death Logs	0.220	-0.400	(0.239)
Baseline Monthly Mortality in Media Reports	0.062	-0.036	(0.051)
Panel C: Baseline Staff Survey Responses			
# Staff Responses	24.870	-6.863	(7.138)
Share Male	0.506	0.013	(0.065)
Age	42.818	1.926	(1.891)
Share White	0.654	-0.020	(0.088)
Share Black	0.135	0.097	(0.073)
Share Hispanic	0.138	-0.052	(0.059)
Horizontal Coordination	0.659	-0.004	(0.029)
Vertical Coordination	0.670	-0.009	(0.041)
Satisfaction	0.690	-0.024	(0.020)
Respect Incarcerated Population	0.737	-0.003	(0.031)
F-Stat		1.167	
Panel D: Baseline Jail Characteristics			
ADP	375.812	-169.677	(117.846)
Avg. Stay Length (Months)	1.124	0.008	(0.276)
Medical FTE	5.689	-1.192	(1.276)
Medical FTE Variety	5.717	-0.481	(0.951)
For-Profit Health Vendor	0.652	-0.047	(0.143)
Previously Accredited	0.109	0.051	(0.088)
Share County Voting Republican	0.448	-0.024	(0.040)
Sheriff is Republican	0.326	0.007	(0.126)
F-Stat		1.389	
F-Stat		1.702	
Observations		46	
Strata FE		✓	

Notes: Data are from the baseline facility survey, baseline staff survey, medical audit, and Census of Jails (2019) (U.S. Department of Justice 2022). Column (1) is the sample mean, column (2) and (3) report the coefficient on an indicator for treatment and the associated robust standard error, respectively. Panel (A) compares baseline quality standards (for detailed definitions of the standards and their calculation see Appendix Section I). Panel (B) compares the baseline mortality in death logs and in media reports. Share of health staff is the number of health staff who took the staff survey divided by the total number of staff that took the staff survey. ADP stands for average daily population. FTE is the number of full-time equivalents of health staff. For-Profit Health Vendor indicates that the facility has hired an external private company to provide health services. Horizontal coordination, horizontal coordination, satisfaction, timely care and respect incarcerated population are staff sentiment indices constructed by individual questions. These questions are asked to health and custody staff and are answered on a Likert scale re-normalized on a scale from 0 to 1 with higher values representing stronger agreement. Joint *F*-statistic is listed at the bottom of the table. "Baseline Monthly Mortality in Death Logs" is excluded from the omnibus test of balance to ensure a complete sample, as death logs are only available for 41 out of 46 facilities. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Table A15: Effect of Accreditation Completion on Staff Sentiment Indices: 2SLS Estimates, Starting Sample

	Horizontal Coordination	Vertical Coordination	Satisfaction	Respect Inmates
	(1)	(2)	(3)	(4)
	Panel A: Second Stage			
$\widehat{Accreditation}$	0.059**	0.037	0.037**	-0.011
	(0.027)	(0.031)	(0.018)	(0.023)
	Panel B: First Stage			
Treat	0.804***	0.779***	0.768***	0.782***
	(0.079)	(0.097)	(0.103)	(0.095)
F-stat	104.39	64.53	55.51	67.90
Control Complier Mean	0.618	0.627	0.639	0.764
Observations	674	674	674	674
Number of Clusters	46	46	46	46
Strata FE	✓	✓	✓	✓
Baseline Control	✓	✓	✓	✓

Notes: This table reports 2SLS estimates of the treatment effect of accreditation on five staff sentiment indices obtained from estimating Equation 2. All specifications include randomization strata fixed effects and the baseline mean of the respective outcome variable. Staff sentiment indices are created using data from the baseline/endline staff survey. The horizontal coordination index includes: (1) "Custody staff support the implementation of clinical decisions and partner with clinical staff"; (2) "Custody and clinical team staff work collaboratively to give good health care to inmates"; (3) "There is open communication between custody and health staff." The vertical coordination index includes: (1) "Staff feel comfortable speaking up when they see something that may negatively affect inmate care"; (2) "When staff speak up about inmate safety and care decisions, those with greater authority are open to their concerns"; (3) "My supervisor considers staff suggestions for improving inmate health and safety"; (4) "My supervisor considers staff suggestions for improving staff health and safety"; (5) "My supervisor takes direct actions to address staff concerns for improving inmate health and safety"; (6) "My supervisor takes direct actions to address staff concerns for improving staff health and safety." Questions (1) and (2) are only asked to health staff. The satisfaction index includes: (1) "I received the training I need to do my job well"; (2) "I have the tools and resources I need to do my job well"; (3) "I am satisfied with my job"; (4) "I find my job to be meaningful"; (5) "I feel valued at my job"; (6) "I find my job to be challenging"; (7) "I would recommend my job to others." The respect inmates index includes: (1) "Inmates at this facility are treated with respect." These questions are answered on a Likert scale renormalized on a scale from 0 to 1 with higher values representing stronger agreement. Control complier mean is computed as described in the main text. We obtained endline responses from 40 jails and we impute missing endline responses for the other 4 facilities in our sample with baseline values. Standard errors are clustered at the facility level. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A16: Effect of Accreditation Completion on Meta Quality Standards: 2SLS Estimates, Starting Sample

	Simple Mean	Principal Component	Expert Weights
	(1)	(2)	(3)
	Panel A: Second Stage		
$\widehat{Accreditation}$	0.041	0.105*	0.072*
	(0.034)	(0.063)	(0.041)
	Panel B: First Stage		
Treat	0.548***	0.545***	0.534***
	(0.087)	(0.090)	(0.093)
F-stat	39.44	36.89	33.19
Control Complier Mean	0.629	0.683	0.606
Observations	46	46	46
Baseline Control	✓	✓	✓
Strata FE	✓	✓	✓

Notes: This table reports 2SLS estimates of the treatment effect of accreditation completion on meta quality standards by combining seven standards weighted by expert survey responses, from estimating Equation 2 (for detailed definitions of the standards and their calculation see Appendix Section I and for expert weights see Figure 6). All specifications include the baseline quality standard and randomization strata fixed effects. The outcome variable is the endline quality standard. Control complier mean is computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A17: Effect of Accreditation Completion on Quality Standards: 2SLS Estimates, Starting Sample

	Gov. Admin. (1)	Safety & Prevention (2)	Personnel Training (3)	Ancillary Services (4)	Patient Care & Rx. (5)	Special Needs (6)	Medical Legal (7)
	Panel A: Second Stage						
$\widehat{Accreditation}$	0.014	0.080	0.133*	0.035	0.090*	0.005	0.106
	(0.031)	(0.057)	(0.070)	(0.032)	(0.048)	(0.078)	(0.078)
	Panel B: First Stage						
Treat	0.575***	0.498***	0.540***	0.543***	0.555***	0.513***	0.539***
	(0.085)	(0.097)	(0.098)	(0.098)	(0.090)	(0.093)	(0.098)
F-stat	46.17	26.46	30.26	30.40	37.92	30.30	30.15
Control Complier Mean	0.626	0.623	0.550	0.559	0.586	0.699	0.669
Observations	46	46	46	46	46	46	46
Baseline Control	✓	✓	✓	✓	✓	✓	✓
Strata FE	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports 2SLS estimates of the treatment effect of accreditation completion on seven quality standards from estimating Equation 2 (for detailed definitions of the standards and their calculation see Appendix Section I). All specifications include the baseline quality standard and randomization strata fixed effects. The outcome variable is the endline quality standard. Control complier mean is computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

Appendix Table A18: Effect of Accreditation Completion on Mortality: 2SLS Estimates Starting Sample

	(1)	(2)	(3)
	Panel A: Second Stage		
$\widehat{Accreditation}$	-0.298**	-0.221***	-0.246**
	(0.119)	(0.075)	(0.096)
	Panel B: First Stage		
Treat	0.537***	0.562***	0.527***
	(0.100)	(0.101)	(0.105)
F-stat	28.61	30.88	25.42
Control Complier Mean	0.337	0.260	0.287
Observations	46	46	46
Strata FE	✓	✓	✓
Control for ADP		✓	
Baseline Control			✓

Notes: This table reports 2SLS estimates of the effect of accreditation completion on mortality obtained from estimating equation 2. All specifications include randomization strata fixed effects. Baseline mortality refers to the average number of deaths per month occurring during the six months preceding treatment assignment. The outcome variable is the average number of deaths per month occurring in the six-month window starting ten months after treatment assignment. Column (2) controls for the log of the current average daily population (ADP) of the jail. Column (3) replaces ADP with baseline mortality. Control complier mean is computed as described in the main text. Robust standard errors are reported in parentheses. $*$, $**$, and $***$ refer to the statistical significance at 10, 5, and 1 percent level, respectively.

B Model of Accreditation

Setup: Consider a principal (the sheriff, S) and two agents indexed by i: custody (C) and medical (M). Agents reduce deaths by simultaneously choosing high ($e_i = h$) or low ($e_i = l$) effort, with $0 < l < h < 1$. We assume agent effort is complementary and costs to effort are linear and are given by the function $C(e_i) = e_i$. The principal incurs legal liability for mortality (d) in excess of the expected mortality rate among the incarcerated population and penalizes agents according to their effort levels.

The observed mortality rate d is given by:

$$d = 1 - \lambda(\min\{e_M, e_C\}), \tag{3}$$

where λ is the marginal product of effort in reducing death and lies on the interval $[0, \frac{1}{h})$.[29] The observed mortality rate is a deterministic function of effort as most deaths are preventable with effort and there are minimal external shocks that affect the mortality rate.

The principal faces liability (L) when the observed mortality rate exceeds a threshold $\hat{d}$, defined as the expected mortality rate when both agents exert maximal effort:

$$L = \begin{cases} 0 & \text{if } d \in [0, \hat{d}] \\ d - \hat{d} & \text{if } d > \hat{d}, \end{cases} \tag{4}$$

such that there is no liability if the observed mortality rate is below the threshold $\hat{d} \equiv 1 - \lambda h$, but there is liability above the threshold which is increasing in the excess mortality rate $(d - \hat{d})$.[30] Since agent's effort level is unobservable to the principal, the marginal penalty η_i is the same: $\eta_m = \eta_c$. The total penalty for agent i is given by:

$$\eta_i[L(d)] = \begin{cases} 0 & \text{if } \tilde{e} = h \\ \eta_i \lambda(h - l) & \text{if } \tilde{e} = l, \end{cases} \tag{5}$$

We assume that $\frac{1}{2}\eta B < (h - l) < \eta B$, with $B = \lambda(h - l)$ implying that the relative effort cost is bounded by the penalty under uncertainty and with certainty.

Information: There are two sources of information friction: (1) agents are unaware of the other agent's type; and (2) the principal cannot observe agent effort directly. We model these frictions as follows:

(1). Hidden Attribute: Agents M and C each have two types: $T_m \in \{t_m^1, t_m^2\}$ and $T_c \in \{t_c^1, t_c^2\}$ giving rise to four different states of the world: $\Omega = \{(t_m^1, t_c^1), (t_m^1, t_c^2), (t_m^2, t_c^1), (t_m^2, t_c^2)\}$. If each state is equally likely then: $p(\omega) = \frac{1}{4}$ for all $\omega \in \Omega$. Agents are aware of their own type but have flat priors over the type of the other agent.[31] In the absence of accreditation, the expected utility of a given effort level given uncertainty on other's type is: $U_i(e_i) = -(\eta_i \mathbb{E}[L(d)] + e_i)$.

(2). Hidden Action: In the absence of accreditation, the principal cannot observe the individual agents' actions, rather he observes the aggregate output (*i.e.*, mortality).

[29] 1 corresponds to the maximum mortality rate.

[30] It is straightforward to allow the excess mortality rate $(d - \hat{d})$ to be multiplied by a monetary scalar to reflect damages.

[31] This is equivalent to receiving an imperfect signal s^1 or s^2 (with equal likelihood), which indicates that the other agent is of type t^1 or t^2, respectively such that $Pr(s^1|t_{-i}^1) = \frac{1}{2}$ and $Pr(s^2|t_{-i}^2) = \frac{1}{2}$. The posterior beliefs are given by Bayes' rule: $p_i'(t_{-i}^1|s^1) = \frac{Pr(s^1|t_{-i}^1)Pr(t_{-i}^1)}{Pr(s^1)} = \frac{1/2.1/2}{1/2} = \frac{1}{2}$.

B.1 No Accreditation

In the absence of accreditation, hidden information is represented by a private payoff θ which takes two values corresponding to agent type and represents an aversion to coordination. Hidden action implies that any penalty levied by the principal is equally split between agents. Below we show that in the baseline case, low effort is the unique Bayesian Nash Equilibrium (BNE):

	$e_C = h$	$e_C = l$
$e_M = h$	$(-h-\theta, -h-\theta)$	$(-\eta B - h, -\eta B - l)$
$e_M = l$	$(-\eta B - l, -\eta B - h)$	$(-\eta B - l, -\eta B - l)$

where, the private payoff $\theta = 0$ if agents are of type t^1 and $\theta > \eta B$ when agents are of type t^2. l is the dominant strategy for type t^2 agents as θ is large enough to make $e_i = l$ the best response for agent i when $e_{-i} = h$. Agent's best response to $e_{-i} = l$ is always $e_i = l$ as $h > l$.

To prove a given set of strategies $(s^*_M(t^1), s^*_M(t^2), s^*_C(t^1), s^*_C(t^2))$ is a BNE, we need to show that each player of each type is playing their optimal strategy given the conditional beliefs of that type against the optimal strategies of all other players. (ll, ll) is the BNE in the absence of accreditation.

We show that (ll, ll) is the BNE in the absence of accreditation as no agent have an incentive to deviate. First, both agents of type t^2 will play l as it is a dominant strategy for that type. Next, we need to check the possible equilibrium strategies for type t^1. We show below that a player of type t_1 will always match the strategy of the other player.

- Agent M's best response to the strategy profile ll of agent C is ll.
  We have shown that agent M of type t^2 will choose l as it is the dominant strategy. Therefore, we need to show that agent M of type t^1 will choose l over h, implying that $U_M(ll, ll|t^1) > U_M(hl, ll|t^1)$.
  Utility from h: $U_M(hl, ll|t^1) = \frac{1}{2}(-\eta B - h) + \frac{1}{2}(-\eta B - h) = (-\eta B - h)$
  Utility from l: $U_M(ll, ll|t^1) = -\eta B - l$
  As $-\eta B - l > -\eta B - h$, l is the best response for type t^1 agent to a ll strategy by agent C.
- As payoffs are symmetric, agent C's best response to the strategy profile ll of agent M is ll.

Next, we prove that (hl, hl) is not a BNE.

- We have shown that agent M of type t^2 has l as a dominant strategy. We need to show that agent M of type t^1 chooses h, when agent C plays (hl), which would require $U_M(hl, hl|t^1) > U_M(ll, hl|t^1)$. $U_M(hl, hl|t^1) = \frac{1}{2}(-h) + \frac{1}{2}(-\eta B - h)$ and $U_M(ll, hl|t^1) = -\eta B - l$. However, $U_M(hl, hl|t^1) < U_M(ll, hl|t^1)$ as $\eta B < 2(h - l)$, which is a contradiction.

B.2 Accreditation as Coordination

Accreditation can resolve the two information frictions leading to coordination on a new, high effort BNE.
a. Horizontal Coordination: Accreditation could serve as a perfect signal of agent type. If $\omega = (t^1_m, t^1_c)$, both agents receive the signal s^1 and update via Bayes' rule resulting in the following posterior beliefs: $p'_{-i}(t^1_{-i}|s^1) = 1$.[32] Now, we can show that in addition to (ll, ll), (hl, hl) is also a BNE.

[32] $p'_{-i}(t^1_{-i}|s^1) = \frac{Pr(s^1|t^1_{-i})Pr(t^1_{-i})}{Pr(s^1)} = \frac{1.1/2}{1/2} = 1$

- Agent M plays h when their type is t^1 and when agent C plays hl.
  Utility from h: $U_M(hl, hl|t^1) = (-h) + 0(-\eta B - h) = -h$
  Utility from l: $U_M(ll, hl|t^1) = -\eta B - l$
  As $-h > -B - l$, h is the best response for type t^1 agent.
- Agent C plays h their type is t^1 and when agent M plays hl.
  Utility from l : $U_C(hl, ll|t^1) = -\eta B - l$
  Utility from h: $U_C(hl, hl|t^1) = (-h) + 0(-\eta B - h) = -h$
  As $-h > -\eta B - l$, h is the best response for type t^1 agent.

Note that both agents of type t^2 will still play l as that is the dominant strategy. Thus, a perfect signal of the state by itself will not result in a unique (hl, hl) BNE.

b. Vertical Coordination: If in addition to each agent knowing the other agent's type, the principal is informed of individual agent actions, then the principal can attribute excess mortality to the specific agent(s) exerting low effort. A large enough specific penalty on the agent choosing low effort allows h to be a dominant strategy for both types and obtain a unique (hh, hh) BNE.

If both agents exert low effort, they continue to receive equal marginal penalty η. However, if one agent exerts low effort while the other agent exerts high effort, the principal can increase the marginal penalty for the agent choosing $e_i = l$ to $\eta_i = \hat{\eta} > \eta$ such that $\hat{\eta}B - (h - l) > \theta$ and reduce the marginal penalty for the agent taking $e_i = h$ to $\eta_i = 0$. When both agents take high effort, $\eta_i = 0$ for both:

	$e_C = h$	$e_C = l$
$e_M = h$	$(-h - \theta, -h - \theta)$	$(-h, -\hat{\eta}B - l)$
$e_M = l$	$(-\hat{\eta}B - l, -h)$	$(-\eta B - l, -\eta B - l)$

Agents of either type will now only choose h as it becomes the dominant strategy.

- An agent of type t^1 chooses h as it is the dominant strategy. If $e_{-i} = l$, the best response of agent i is h as $h - l < \eta B$. Similarly, if $e_{-i} = h$, agent i's best response is h as $h - l < \hat{\eta}B$.
- An agent of type t^2 chooses h as it is the dominant strategy. If $e_{-i} = l$, the best response of agent i is h as $h - l < \eta B$. Similarly, if $e_{-i} = h$, agent i's best response is h as $h - l + \theta < \hat{\eta}B$.

This eliminates the (ll, ll) BNE and results in a unique (hh, hh) BNE as both players of both types always choose h.

2. **Rubber Stamp**: Under the rubber stamp scenario, we assume that $\eta_i = 0$ such that no penalty is levied for whichever action an agent chooses. In such a scenario, (ll, ll) is the unique equilibrium. In a richer model of social welfare, firm profits (for the accrediting agency) might rise and the risk of legal action (against the sheriff) may fall at the cost of taxpayer subsidies for accreditation. The modified payoff matrix is given below:

	$e_C = h$	$e_C = l$
$e_M = h$	$(-h - \theta, -h - \theta)$	$(-h, -l)$
$e_M = l$	$(-l, -h)$	$(-l, -l)$

It is straightforward to see that l becomes the dominant strategy for agents of type t^1 as $-h < -l$.

3. **Distraction**: Accreditation might also serve as a distraction by targeting the wrong inputs and processes.

In this case, choosing h does not reduce mortality any more than choosing l, and therefore the same penalty is levied on agents for both effort levels.

	$e_C = h$	$e_C = l$
$e_M = h$	$(-\eta B - h - \theta, -\eta B - h - \theta)$	$(-\eta B - h, -\eta B - l)$
$e_M = l$	$(-\eta B - l, -\eta B - h)$	$(-\eta B - l, -\eta B - l)$

Despite receiving a perfect signal about the other agent's type, a unique (ll, ll) BNE is obtained and accreditation fails to introduce the (hl, hl) BNE as l becomes a dominant strategy for both types of both agents.

C Recruitment

From July 2021 to April 2022, medium-sized jails across the U.S. were recruited to participate. The study was promoted at targeted meetings and conferences, including those of the Committee of State Sheriffs, the National Sheriffs' Association, the American Jail Association, the National Association of Counties, the National Association of Counties, and NCCHC itself.

In addition, we used the Safety Source Directory 2021 to make random cold calls to recruit additional facilities. The cold call script is available here. Interested facilities could also directly reach out to the Harvard research team.

If a facility expressed interest in the study, the Project Coordinator scheduled a half-hour informational meeting on Zoom with the facility and their medical leadership. During this session, the Project Coordinator provided a short presentation on the study background and components, as well as the potential benefits of NCCHC accreditation. Facilities also had the opportunity to ask questions about the study and accreditation process during these informational meetings.

D Enrollment

Following the informational meeting, the Project Coordinator sent the facility an email summarizing the session and a link to a screener survey used to confirm their eligibility. Criteria for eligibility included: (a) an average daily population (ADP) of 100 to 3,000, and (b) no prior accreditation with the NCCHC, although they could have received a different accreditation process in the past. Facilities that met this criteria received enrollment documents including: (1) a consent form (co-signed by medical and custody); and, (2) a Data Use Agreement (DUA). Facilities were also able to opt in to receive a $500 payment for completing each of the baseline and endline surveys. The study team also applied and was granted a Certificate of Confidentiality (COC) from the Department of Health and Human Services which guaranteed protection of information obtained from all participating facilities in our study, including during federal, state, and local civil/criminal/administrative/legislative proceedings.

After signing the Consent and DUA, the facility was randomly assigned a date to attend an Orientation Meeting. During this meeting, the team answered remaining questions, obtained facility contact information, confirmed prior accreditation status and current health care vendor, and walked through the next steps for the study. The presentation is linked here.

For enrollment into sub-facility surveys, including the inmate interview, leader interview, and staff surveys, individuals within the enrolled facilities were recruited to participate in three different ways. For the inmate interviews, we asked each facility to randomly select three inmates to participate. For the interviews with facility leadership, we asked the facility to select one person on the custody side and one person on the health care side to participate. For the staff survey, facilities provided us with a list of health and custody email addresses to invite to participate in the survey.

E Randomization and Stratification

Facilities were randomized into the treatment and control group after completing the baseline survey to minimize attrition. Facilities randomized to the treatment group (Accredit Now) began the accreditation process immediately, except for three facilities that attrited shortly after receiving the assignment (see Figure 3). Because many correctional facilities cited monetary costs as a major barrier to seeking accreditation, the Accredit Now facilities accreditation costs were highly subsidized. Facilities randomized to the control group (Accredit Later) were invited to undergo the accreditation process after completing the endline survey (see Figure 2) for details of study and accreditation process timeline).

Randomization occurred in three batches to avoid asking the facilities that were recruited earlier to wait before receiving their randomization status, thus minimizing the likelihood of attrition. Within each batch, we also stratified facilities into two groups based on average daily population (ADP), given the correlation between facility size and health outcomes in the literature. Facilities whose ADP was larger than the median ADP of the batch were considered "large" while facilities whose ADP was smaller than the median were considered "small."

F Accreditation Process Initiation and Completion

We anticipated, given the length of the study and complexity of working at the intersection of the criminal justice and health care systems, there would be unavoidable disruptions that would inhibit facilities from participating in the process fully. So, we continued to recruit and built up a waitlist of facilities to randomize in the event of such disruptions per our pre-analysis plan. Over the course of the study period, two "Accredit Now" facilities did not begin the accreditation process for two different reasons:

1. In the first facility the consenting partner had a heart attack.
2. In the second facility the consenting partner was sued for sexual harassment.

In total, six facilities were recruited later in the study period to replace and re-balance the sample. This resulted in a total number of 46 jails ever randomized and 44 are used in our main analysis since accreditation does not predict either initiation or completion of materials in the process

We find that results are indistinguishable between the main and starting sample using imputation methods for two involuntary removal facilities (see Appendix Tables A14, A15, A16, and A18) and failure to initiate or complete various steps in the process is not predicted by treatment status (see Appendix Table A3).

G Description of Data Elements

G.1 Facility Survey

After the orientation meeting, the facility received the link to the baseline facility survey. Each facility had 2 weeks to complete it and received email and phone reminders 5 days, 3 days, and 1 day before the survey deadlines.

At endline, the facility was asked to update their answers to the survey. The survey they received was prepopulated with their answers from baseline and they only had to update the answers to the questions that had changed, in addition to answering the three new questions that were added to the endline survey. The three new questions added were as follows: (1) A copy of inmate booking records from 2021 to present (including the name of individual, DOB, gender, race, and arrest charges); (2) Number of health care related lawsuits and/or consent decrees filed against the facility by year from 2017 to present; and (3) Estimate of the amount spent on health care each year from 2017 to present. The endline facility survey is available here.

After the facility submitted the online facility survey at both baseline and endline, the research team reviewed their responses and followed up with any clarifying questions regarding information provided or missing documents. Facilities had 14 days to provide missing information or clarify questions. If the information was not provided within the 14 days, the clarification request was closed and research team used the facility's original survey response. If a facility replied with their information late, the team evaluated whether it should be included in the final version of the survey response for coding. See the clarification request email linked here.

The material uploaded by the facility to the online survey was then downloaded to a secure server in a facility-specific folder. The research team then performed initial cleaning of the data and prepared documents for in-depth coding.

G.2 Staff Survey

After the facility survey was sent, the research team obtained staff email lists from each facility. This list included both custody and medical staff emails. The template email is available here and the staff survey questions are linked here. These questions are related to staff satisfaction, coordination, and perceptions of care. In some cases, the facility was unable to provide the list of emails, and we then worked directly with our facility contacts who distributed these surveys on our behalf. The deadline to complete the staff survey was two weeks.

The study team sent reminder emails 5 days, 3 days, and 1 day before the survey deadline to increase the number of responses. If at least 5 individuals completed the survey and a gift card lottery was permitted by the facility, one staff member was randomly selected to receive a $100 gift card.

At endline, we requested an updated staff email list and sent the survey to both the baseline and endline email list.

G.3 Medical Audit

Each facility completed a virtual medical audit at endline. To be able to measure the effect of accreditation and changes over time, medical charts were requested for both the endline period and period of time before

baseline. Each facility received a checklist of material to prepare in advance of the audit available here which included health records for those with suicidal ideation, hypertension, and opioid use disorder, if available, as well as two inmates at random. We also collected the 12-month death log, sick log and any administrative reviews of death or autopsies.

Each facility was able to select a three-hour timeslot for the audit to occur. The medical audit was conducted by a medical expert who was blind to the treatment status of a facility and led the completion of a catalog of standardized questions available here. If the facility used an electronic medical record system, the facility screen shared the records live on the call with the team. If the facility used paper records, the facility received a document camera from the Harvard study team to use to display the records virtually. In all cases, the identifying information of the patient was obscured and not recorded by the study team.

G.4 Leadership Interviews

At endline, qualitative interviews were conducted with one health leader (*e.g.*, nurse of health services administrator) and one custody leader (*e.g.*, warden or sheriff) at each facility to understand their perceptions of the health care delivery. These interviews were conducted on Zoom, took approximately 45 minutes each, and were conducted in private. The interview script, including questions, is available here.

G.5 Inmate Interviews

At endline, qualitative interviews were also conducted with three incarcerated individuals at each facility. The facility was asked to randomly select three incarcerated individuals to participate, of whom at least two should have received health care in the last month. The goal of these interviews was to understand their perceptions of the health care. These interviews were conducted on Zoom, took approximately 45 minutes each, and were conducted in private. The interview script, including questions, is available here.

G.6 Expert Prediction Survey

Definition of Expertise: We recruited respondents with different *types* of expertise to a prediction survey. Given that our focus was on health services in jails, the three domains of expertise were considered: (1) Jail administrators and healthcare providers, (2) Previously incarcerated individuals, and (3) Policy analysts and academics.

Recruitment to Prediction Survey: Jail Administrators and jail health care providers were recruited to complete the survey at national corrections conferences. In addition, we used Safety Source (2021) to cold call various directors of facilities that were not associated with the study. Previously incarcerated individuals were recruited to participate through a screener survey posted on Prolific. To be eligible for the survey, the screener asked them to select all statements that apply to them from a list of 13 options with one of the options being that they have personal experience with the criminal justice system. If they select this option, they then receive another questions asking to select which experience align to their experience. All respondents that selected that they have previously spent time in jail, were then invited to take the expert survey. Finally, policy analysts and academics were recruited at a variety of conferences.

Expert Prediction Survey Aims: The purpose of the expert surveys was to: (1) assess whether accreditation was predicted to affect outcomes and the directionality of that effect; (2) assess whether the effect differed

by type of expertise; (3) ascertain baseline features of facilities might be important for heterogeneity; and (4) compare expert predictions to the study findings.

Deployment. The survey was deployed through Qualtrics (available here).

Incentives All participants received a $20 gift card for participation. To incentivize survey accuracy, three experts and two previously incarcerated individuals with survey responses most accurately aligned with the results of the study are to receive an additional $100 gift card.

Survey Flow and Demographics Following the collection of basic individual background characteristics and a brief description of the study, participants were asked what baseline factors might modify the effects of accreditation. Summary statistics on the experts can be found in Appendix Table G1. Next, respondents were asked *whether* accreditation would affect patient care and treatment, inmate deaths, litigation and health care related spending. We incentivized responses to such answers noting that those who were closest to actual results would recieve a bonus payment. In addition, we asked participants to rank the importance of the quality standards and define what is included in those categories. For many of these questions, we also asked the degree of confidence respondents had in the answers. We use these data to understand whether experts from different domains ranked outcomes differently suggesting they value or weight things differently and to take a disciplined approach to assessing for heterogeneity.

Appendix Table G1: Expert Survey Demographics

	Number of Respondents	Share of Respondents
Age		
18-24	3	2.07
25-34	23	15.86
35-44	41	28.28
45-54	51	35.17
55-64	20	13.79
65-74	5	3.45
75+	2	1.38
Gender		
Male	66	45.52
Female	78	53.79
Transgender	1	0.69
Race/Ethnicity		
Asian	12	8.28
Black or African	19	13.10
Hispanic	4	2.76
Native American	1	0.69
Mixed race	6	4.14
White/Caucasian	101	69.66
Prefer not to say	2	1.38
Schooling		
Some college but no degree	22	15.17
Associate's degree (2 years)	22	15.17
Bachelor's degree (4 years)	29	20.00
Post-graduate degree	64	44.14
High school degree or less	8	5.52

Notes: This table shows the demographics of those who participated in the expert survey. The number of recruited experts is 145. The first column lists the key demographic information, the second is the number of participants that fall into that group, and the third column calculates the share of respondents within that group (adding to 100 for each).

H Coding

Coding for this project required two major tasks: (1) Coding up raw material (*e.g.*, responses and forms from facilities); and (2) Data cleaning and analysis. We also developed processes and procedures to make sure coding was performed in a systematic, unbiased way across various members of the team.

H.1 Coding Committees and Codebook

One of the more laborious tasks was translating the NCCHC questionnaires into analyzable data elements. This required determining which questions were most objective and thus should be included in our Harvard study. We also had to develop a mapping between the questions and binary fields for whether that particular component of the standard was met or unmet (Yes/No, 1/0).

Coding Committees were assembled who worked for several months examining each individual NCCHC question and developing the most objective version of that question for Harvard surveyors as well as specific guidance as to whether the particular criteria was met. The coding committee for the Facility Survey consisted of Harvard undergraduates working with the study authors and Coordinator, alongside predoctral fellows. The Coding Committee for the medical audits included medical students, the Project Coordinator and was overseen by one of the PIs with expertise in medicine. The Facility Survey Committee created a codebook for the coding by research assistants, who typically were Harvard college or Harvard law students. The codebook described how each answer should be coded as to ensure consistency across RAs. Importantly, the RAs were trained to code *whether the service/process* was done as indicative of meeting the standard, not the *outcome of that test/process itself*. For example, *testing* for sexually transmitted infections (STI) was considered meeting the standard even if the patient was ultimately diagnosed with a STI. Each row of the codebook defined the variable name, format (*e.g.*, binary, text), reasons for skipping, question, instructions for coding, valid entries, and the variable description. A link to the codebook is available here.[33] See Appendix Section H.4 for further details.

H.2 Variable Transformation

The last step was the more familiar data cleaning and analysis by predoctoral fellows from the Health Inequality Lab and the Malcolm Wiener Center. This required resolving disagreements between two independent manual coding RAs with a third tie-breaker if needed. Transformation of variables was also required. Variables were recoded so that higher values were preferred. In addition, we performed the following transformations:

1.) For binary questions – we maintained the dichotomous structure and applied the penalty rule described below.

2.) For continuous outcomes – we dichotomized them for purposes of aggregation using the NCCHC standard rule whenever it was specified or the median across the baseline sample.

   a.) For example, for the question how many days after admission was a mental health assessment performed, NCCHC provides a clear standard (within 14 days) and any facility not meeting that standard was given a 0.

[33] At endline, only the answers that changed from the baseline survey were coded.

b.) As an example of the second approach, for the question, "how many FTE does a mental health specialist contribute?" - NCCHC does not have a clear standard. Thus higher than median at baseline was given the value of 1 and lower than median was given a value of 0.

3.) For grouped outcomes, we use the number of outcomes in the group as the denominator and compute the share of all items in the group that were performed as the numerator. For example, one of the grouped questions was whether health records documented patient education on salt intake, weight management or medication.

4.) For some text variables, we encoded to a binary variable. For example, to the question "What is the name of the Responsible Health Authority (RHA)" we coded as 0 if there was no name given as that indicated there was no RHA.

In addition, non-response was often indicative that standards were not being met and we developed specific procedures for coding missings (see Appendix Section H.2.2).

H.2.1 Training of the Manual Coding Research Assistants

In order to be hired as a manual coding Research Assistant for the project, students had to undergo several steps. First, the applicants had to sign a form stating they would maintain strict confidentiality if they happened upon identifying information about a facility in the process of coding up their responses. Second, they did CITI training – which was required by the IRB. Third, they had to complete a training class with the Project Coordinator. The training guided new coders through the process, showed them where all the documents were located, and what information they contained. Lastly, they completed a coding test which was supervised by a senior coder consisting of a spreadsheet with each of the 202 variables across the top. The coding was checked for accuracy and feedback was provided. A particular question was manipulated to assess whether the coder would detect the ambiguous response and reach out to study leadership. If the coder passed these checks, they were used in the actual manual coding exercise.

H.2.2 Penalization for Missing Data

If a facility was unable or unwilling to provide information we assume the worst case scenario.

1. **Refused participation in survey**: If a facility refused to conduct a crucial quantitative part of the Harvard study – either a medical audit or the endline facility survey – all responses are coded as unmet (in the case of medical audit) or unchanged (in the case of refusing to answer an endline facility data).
2. **Refused to answer a specific question**: If a facility participated in the survey but refused to answer a specific question, the response is coded as not meeting the specific criteria.
3. **Gateway questions**: Whether some questions are asked depends on how a previous question is answered. For example, during the medical audit, if a facility answers Yes to "Is the inmate tested for STIs," they are then asked "Of the sample tested for STIs, which diseases were tested: Chlamydia, Syphilis, HIV."" However, facilities that answer no to "Is the inmate tested for STIs," are not asked the following questions, and were penalized as not having met any of the subsequent criteria.

H.3 Coding Outcomes

An important coding decision was how to reduce the dimensionality of the outcomes in a principled manner. This was done by categorizing all the various questions in the facility survey and medical audit into one of the seven quality standards. Given the importance of mortality as an outcome, it was coded separately. See Appendix Section I for further details.

H.4 Coding Facility Survey

The facility survey consisted of 58 questions that were asked directly to a jail administrator and 42 requested documents. Based on the work of the *Coding Committee* the responses were divided two types of variables: (1) coder to determine (CTD) in which RAs determined whether the each answer provided by the facility met the corresponding quality metric and manually coded this determination, and (2) facility to answer (FTA) in which the facility administrator answered directly through the online portal. Each of the FTA questions were then mapped to the seven quality standards and to 40 substandards. Of the 328 CTD questions, the codebook specifically provided instructions on how to ascertain whether the criteria was being met. The coding process for the Facility Survey was as follows:

1. For each facility, two RAs independently coded the baseline facility survey by inspecting the uploaded documentation and following the codebook.
2. To ensure independence and to maintain security, each RA received a personal folder only they could access on the server – and with only the one facility's information they were coding at the time. This was done to prevent RAs from checking their work against others and to preserver security by keeping data access strictly limited.
3. The two manual coding RAs were blinded as to whether the facility was in the treatment or control groups.
4. After completion, a third team member resolved any flagged discrepancies by re-reviewing the entire section the discrepancy was found (not just the question item). The third team member was also blind to the first two reviewers coding.

H.5 Coding Staff Survey

These data are individual-level data with most of the answers being multiple choice. For Likert outcomes, we rescaled so the outcomes were between 0 and 1 but kept continuous.

H.6 Coding Medical Audits

For the medical audits, we received raw workbooks from the medical scribes, which we then cleaned and adjusted certain variables based on NCCHC standards. For example, if an incarcerated individual had not been in the facility for at least 14 days, it would not be applicable to have an initial health assessment. Additionally, when specific medical items (e.g., STI testing, receiving screening) are not recorded, we penalize them by coding them as zeros, indicating that the facility did not meet the standard. These data are inmate-patient level, so patient records are collapsed to the facility level to create a facility-level average.

H.7 Coding Litigation

For litigation, we trained law students to scrape and code information on lawsuits using case filings and decisions from LexisNexis, CourtListener and PacerMonitor. The coding process is similar to the coding of the facility survey as detailed in Appendix Section H.4. A link to the coding instruction is available here. For each lawsuit, coders are asked to identify incident date(s), filing date, disposition, and subject matter of claims. For example, claims might be related to lack of healthcare, condition of confinement, excessive use of force, etc.

I Construction of Indices

As per our pre-analysis plan, we created indices corresponding to each of the quality standards. The indices were created using the questions within each of the data sources (facility survey, both CTD and FTA as well as medical audit) that corresponded to that standard. The NCCHC groups their standards into seven main categories, listed below and described in further detail in Table 1 with typical examples:

- Governance and Administration
- Health Promotion, Safety, and Disease Prevention
- Personnel and Training
- Ancillary health care Services
- Patient Care and Treatment
- Special Needs and Services
- Medical-Legal Issues

To construct the indices we aggregate questions that pertain to a given standard. We aggregate them using the pre-specified the approaches described below:

Equal Weights: Each question was equal weight by computing the simple average across all variables in a given index.

Principal Component Analysis: We also use principal component analysis (PCA) to aggregate individual components (Bai 2003). We include all 214 outcomes and use the first component to create a meta-index, rescaling to ensure the principal component is positive and lies in the unit interval (Lise and Postel-Vinay 2020).

Expert Weights: We created a weighted overall average with the weights corresponding to the importance of each standard as ranked by experts (See Appendix Section G.6 for further details).[34]

[34] This method is also inspired by work from Pew Charitable Trusts (2017). We computed weights from reciprocal ranks as described in Section V.4.

J Cost-Benefit Analysis

Table J1 below presents our relevant RF estimates and lower and upper bounds on each associated benefit and cost of accreditation.

J.1 Benefits of Accreditation

Mortality: Our RF estimates indicate that receiving accreditation reduces monthly mortality rate by 0.14, which is equivalent to 1.7 fewer deaths per year. Given a value of statistical life of $13.2 million (DOT 2024), a 95% CI of the benefit from reducing mortality ranges from $5,879,808 to $38,788,992.

Litigation: Our RF estimates indicate that receiving accreditation results in 0.129 fewer lawsuits over six months, corresponding to 0.258 fewer lawsuits per year. We construct a weighted average jury award of lawsuits from cases alleging "physical illness," "drug/alcohol withdrawal," "overdose," "suicide," and "assault by another incarcerated person" from El-Sabawi et al. (2023). The average weighted litigation cost per case is $1,101,786.1. Thus, a 95% CI of the reduction in litigation costs ranges from -$60,863 to $345,123.

Recidivism: Our RF estimates indicate individuals released from accredited facilities are 12.9% percentage points less likely to be rebooked six months post-release. We take a conservative approach by using the six-month recidivism rate as an annual measure. We multiple this estimate by the average number of annual admissions for the accredited facilities and calculate a social cost weighted crime cost based on the distribution of crime types which comes from Durose and Antenangeli (2021). Crime types include "violent," "property," "drug," and "public order," and estimates of crime costs for each type comes from Miller et al. (2021). The average weighted cost of future crime is $20,690 per rearrest. This results in a 95% CI of benefits ranging from $1,577,750 to $21,978,932.

J.2 Costs of Accreditation

Accreditation Fee: The cost of accreditation includes the initial accreditation fee plus the focus survey fee for facilities (if needed). The average accreditation cost for the accredited facilities is $8,199.

Personnel Cost: We take a conservative approach and assign a personnel cost associated with obtaining accreditation based on the total FTE of staff at the accredited facilities, multiplied by the average share of the year spent on the accreditation process of 0.63. Using the Census of Jail (U.S. Department of Justice 2022), the average number of health staff and custody staff in the accredited facilities is 27 and 106, respectively. Based on our online scraping, the average annual salary for health and custody staff scraped is $81,266 and $49,169, respectively. Taken together, the personnel cost is $4,665,841.

J.3 Cost-Benefit Calculation

We combine our estimates of the benefits and costs of receiving accreditation. We estimate a lower-bound net benefit of accreditation over one year of $2,722,655. The upper-bound net benefit is =$56,439,007.

Appendix Table J1: Cost-Benefit Calculation

	RF Estimates	Lower Bound	Upper Bound
	(1)	(2)	(3)
Panel A: Social Benefits of Accreditation			
Mortality	-0.141	$5,879,808	$38,788,992
	(0.053)		
Litigation	-0.129	-$60,863	$345,123
	(0.094)		
Recidivism	-0.129	$1,577,750	$21,978,932
	(0.057)		
Panel B: Social Costs of Accreditation			
Accreditation Fee	-	$8,199	
	-		
Personnel Cost (Health)	-	$1,382,335	
	-		
Personnel Cost (Custody)	-	$3,283,506	
	-		

Notes: This table shows the cost benefit analysis. Panel A shows the benefits of accreditation. RF estimate for mortality comes from Column (3) of Table A11. RF estimate for litigation comes from Column (6) of Table A13. RF estimate for recidivism comes from Column (3) of Table A13.

K Deviations from Pre-Analysis Plan

For the most part, we followed our pre-analysis plan (Banerjee et al. 2020), but there were a handful of deviations which were driven by the logistical exigences from the field, and are describe below.

1. We had considered using both a quality standards index and a production function index (the latter we had devised). However, we ended up focusing attention on the NCCHC quality standards index since these were more readily translatable to policymakers.
2. We expected to use the provision of death logs as a quality measure, but realized that death log data also contain information on deaths (number and cause), allowing us to assess whether mortality is affected directly.
3. In addition to using intent-to-treat, we use instrumental variables. We did not anticipate that a substantial share of facilities would fail to complete the accreditation process.
4. Regarding the timing of endline, we had originally planned for endline to begin 14 months after baseline. However, this was not possible due to delayed scheduling of the on-site NCCHC visits. Instead, facilities are broken into three main batches for endline: (1) Spring-Summer, (2) Summer-Fall, and (3) Late Enroll. The endline process began in May 2023 for the Spring-Summer batch, in July 2023 for the Summer-Fall batch, and in March 2024 for the Late Enroll batch. These windows are created by taking the minimum and maximum of the latest and earliest dates of endline for Accredit Now facilities within each strata. We note, however, that some facilities were still not finished by the endline given the iterative process that shown in Figure 1.
5. We anticipated, given the length of the study, disruptions that would be orthogonal to treatment status, so we continued to recruit and built up a waitlist of facilities to randomize in the event of such events. Over the course of the study period, two "Accredit Now" facilities did exit immediately after receiving their treatment assignment. In one facility, the consenting party had a heart attack and was not able to return to the job. In the second case, the consenting party was accused of sexual assault and placed on leave. In both instances, there was no alternate available to be a party to the consent. In total, six facilities were recruited later in the study period due to these involuntary removals. We refer to the 46 facilities as the starting sample with involuntary removals and the 44 jails exclusive of these two medical-legal events as the final analytical sample.

L Institutional Appendix

L.1 Jails

Jails serve as the entry point to the criminal justice system. Though the majority of individuals in jail are awaiting trial, some serve their entire sentences in jail. Each day, there are an average of 652,500 individuals incarcerated in our nation's jails (Zeng 2023).

County sheriffs are responsible for operating their county's jail. Sheriffs are generally elected through popular vote with a few exceptions, such as Rhode Island where the sheriff is appointed (National Sheriffs' Assocation 2015). Sheriffs also do not exist in Connecticut, Hawaii, and Alaska. Term lengths vary by state, and only a handful of states have term limits (Reaves and Hickman 2002). Often, part of the platform for what the sheriff runs on during elections involves custodianship ofs the jails.

L.2 Quality Standards in Jails

As courts fleshed out the constitutional right to adequate health care for incarcerated individuals, various medical and legal organizations joined forces to study the jail landscape and develop the precursors for modern quality standards. In the early 1970s, the American Medical Association (AMA) and the Commission on Correctional Facilities at the American Bar Association (ABA) entered into a joint effort to "institute and improve the medical and health services in the nation's jails and prisons" (Carolynn and Gus 1973). As part of this cooperative agreement, the AMA and ABA performed the first systematic survey of the state of health care in U.S. jails in 1972 (Carolynn and Gus 1973). The results demonstrated the dire situation within jails, with the AMA and ABA concluding that "medical care is virtually nonexistent in scores of U.S. jails and is barely adequate in hundreds more..." Among the 1,159 jails who responded, approximately 70% of jails had only first aid available and another 17% had no medical facilities available at all. Jails primarily relied on community medical facilities, such as local government and private hospitals, to provide care. Indeed, medical personnel within jails was scarce. Less than 40% of responding jails had physicians available on a regularly scheduled basis, with nearly 80% of jails having no formal arrangements with physicians to provide care. These findings, among others resulted in a 1974 joint report by the ABA and AMA on the state of health care in jails and prisons that illustrated the deficiencies in the system(Association 1974).

Indeed, case studies conducted by the ABA, AMA, and states revealed substantial conflicts between medical and custody staff that prevented the delivery of health care. A 1972 report on the health care and conditions in Pennsylvania's state prisons, for example, concluded that "civilian health personnel continually face a dilemma of choosing between treatment and concerns for security" (Association 1974). As a result, access to care is often denied because "[w]hen ill, the prisoner's point of contact with a prison's health care program is the sick-call line. Access may be barred by a guard, who refuses to give the convict a hospital pass out of whimsy or prejudice, or in light of a history of undiagnosed complaints"(Association 1974).

In 1979, the AMA published a first set of national standards for health care in jails (American Medical Association 1979), which reflected substantial input from sheriffs, jail administrators, and health care providers in jails. Recognizing that many jails were facing liability risk following the *Estelle* ruling, these 1979 standards provided some guidance to jails and were designed to "reflect the viewpoint of organized medicine regarding the definition of adequate medical care and health services as they exist in the community, insisted upon by the courts." Critical to these standards was cooperation between medical and custody

staff. Notably, the AMA stated that the "Standards call for close cooperation between the medical staff, other professional staff, correctional personnel and facility administration."[35] By 1983, the program moved outside of the AMA with funding from the Robert Wood Johnson Foundation and became known as the National Commission on Correctional Health Care (NCCHC), a not-for-profit organization based in Illinois (Gibson and Phillips 2016; National Commission on Correctional Health Care 2018).

L.3 NCCHC

The National Commission on Correctional Health Care (NCCHC) is one of the oldest correctional health care accrediting bodies in the U.S. The NCCHC grew from a program of the American Medical Association (AMA) in the early 1970s after AMA surveys identified disorganization in the provision of health care in jails (Gibson and Phillips 2016). This led to the creation of standards for care and a pilot accreditation program funded by the Law Enforcement Assistance Administration (Modlin 1979). The pilot program slowly expanded and in the 1980s, funding from the Robert Wood Johnson Foundation allowed for the NCCHC to be established as a stand-alone nonprofit (National Commission on Correctional Health Care 2018). Today, NCCHC develops standards of care for jails, prisons, juvenile detention facilities, mental health services, and opioid treatment programs.

Accreditation of jails is based on the NCCHC Standards for Health Services in Jails. This set of consensus-driven standards covers the following areas: governance and administration; health promotion, safety, and disease prevention; personnel and training; ancillary health care services; patient care and treatment; special needs and services; and medical-legal issues.

The NCCHC accreditation process begins with the facility completing a detailed survey which NCCHC uses to identify deficiencies. Next, trained correctional health care experts visit the facility to verify compliance with the standards. The team tours the facility, reviews health records and other documents, and gathers data through observation and interviews with staff and inmates. The lead surveyor submits a report to NCCHC's Accreditation Committee after the visit. This Committee, composed of health professionals, is blinded to the facility's identity and determines compliance based on the report. If the Committee decides the facility is not compliant, NCCHC continues to work with the facility through a process dubbed "corrective action."

To obtain accreditation, a facility must comply with 100% of the "essential standards" and 85% of the "important standards." If a facility is in partial compliance with most of the "essential standards," the facility is placed in "Continuing Accreditation with Verification." These facilities have four months to complete corrective action and will then receive accreditation. If a facility has apparent deficiencies requiring resolution, they will be placed in "Provisional Accreditation." These facilities have four months to complete corrective action and may then be placed in "Continuing Accreditation with Verification," Accredited, or on probation. Facilities that remain non-compliant to many standards, do not submit corrective action in a timely manner, fail to resolve compliance findings, refuse to respond to NCCHC follow-up requests, or falsify information, are placed on "probation." Finally, facilities with significant deficiencies in multiple standards and who remain non-compliant are denied accreditation, as was the case for several facilities in this study.

[35]The AMA also established a set of standards for health services in prisons which provided that medical personnel and other prison officials are to act in "close cooperation and coordination" in a "joint effort."

L.4 Legislation

There is no federal legislation mandating accreditation for health care facilities in the U.S. However, health care systems for the non-incarcerated individuals, are incentivized to be accredited in order to be reimbursed by the Centers for Medicare and Medicaid Services (The Joint Commission 2018). Accreditation authorities such as the Joint Commission provide deemed status which enables clinics and hospitals to receive tax-financed reimbursement.

This is due to the Medicaid Inmate Exclusion Policy which prohibits the use of federal Medicaid funds for inmates unless they are receiving inpatient care in a medical institution such as a hospital for a minimum number of nights (Social Security Amendments 1965).

While this legislation restricts federal taxpayer funds from flowing to incarcerated individuals, counties are still mandated to provide locally taxpayer-provided funds to incarcerated individuals. This is because incarcerated individuals have a constitutional right to health care as ruled by the 1976 Supreme Court case, *Estelle v. Gamble* (*1976*). This landmark Supreme Court case ruled that deliberate failure to provide incarcerated individuals with adequate health care constituted "cruel and unusual punishment" which is prohibited by the Eighth Amendment.

Building on the *Estelle* decision, *Helling v. McKinney* (1993) rules that the Eighth Amendment also protects inmates from "future harm", including the long-term effects of exposure to second-hand smoke. However, the Court also stated that to prove this, evidence must be presented that illustrates: (1) the seriousness and likelihood of harm; and, (2) deliberate intent. These parameters remain the norm today (Alsan et al. 2023). The Prisoner Litigation Reform Act (PLRA) of 1997 places several additional steps and obstacles (*e.g.*, following sometimes multi-step and arduous grievance process within a short window, paying filing fees) for individuals to file suit while incarcerated.

L.5 Evidence on Accreditation and Health

Greenfield and Braithwaite (2008) performed an analysis of existing empirical studies in English on accreditation of health care services in 2007. Across 66 studies that met criteria, the effect of accreditation was "complex." Accreditation was consistently associated with promoting change broadly and professional development – however there were inconsistent effects on measures of quality. A more recent analysis by Mumford et al. (2013) which looked at English language studies on the economic impact of accreditation of health services in 2011. The authors did not use meta-analysis techniques due to the variability in the outcomes. Data on costs and benefits were available for six of fifteen studies. The authors found that benefits of accreditation were "inconclusive." Given these inconclusive and mixed findings of accreditation of health care services broadly, it seemed reasonable to design a RCT. We are aware of only one thus far in the literature – a project in 2009 in South Africa (Salmon et al. 2003) used a prospective randomized trial to assess whether accreditation of public hospitals in Kwazulu Natal affected the standards identified by the organizing agency as well as independent quality measures. This study found that accreditation improved quality metrics, however, the sample size was only 20 hospitals.

According to Sechrest (1976) – the idea of an accreditation program in corrections arose out of a "genuine concern from the field" (*p.* 15). The idea was that better care and services to the incarcerated would lead to long-term social savings, greater public safety and induce a virtuous cycle of increasing support for public funds dedicated to rehabilitation in jails.

A retrospective by Stanley (2009) – a former director of accreditation of NCCHC with over 11 years of experience, concludes that accreditation helps uncover problems, which then custody and health can try and resolve. She notes that accreditation is also about having clear expectations that everyone understands and working towards those goals together. She also states that accreditation not only helps drive better quality health care, but can even protect against lawsuits, create cost-saving efficiencies, and increase coordination and cooperation between health and custody.

On the other hand, Rich, Allen and Williams (2015) outlines the need for higher standards in correctional health care to improve public health. In particular, the article points to enhanced screening and treatment for Hepatitis C, better provision of mental health care and attention to problems of aging behind bars. More generally, there is a need to integrate medical care inside and outside the correctional facility.

M Additional Materials

M.1 Index of Items

Appendix Table M1: Index of Items

Hyperlink to Item	Description of Item
Pre-Analysis Plan	Pre-Analysis Plan.
Facility Survey	The Qualtrics survey facilities completed at baseline and endline.
Facility Survey Codebook	This is the codebook that research assistants used to code-up facility survey responses.
Facility Survey Clarification Email	This email was sent to facilities after study team members reviewed the facility survey to request additional documents or ask questions related to their answers.
Staff Survey	The Qualtircs staff survey that facility staff members were invited to complete at baseline and endline.
Staff Survey Email	This email was sent to health and custody staff members at each facility to invite them to take the staff survey.
Facility Leadership Interview Questionnaire and Script	The questions asked to health and custody leaders as part of the leadership interviews at endline.
Inmate Interview Questionnaire and Script	The questions asked to inmates as part of the inmates interviews at endline.
Expert Survey Prolific Screener Survey	The screening survey posted to Prolific to find individuals who had previously been incarcerated and invite them to complete the full expert survey.
Expert Survey	The Qualtrics expert survey that individuals were invited to complete.
Medical Audit Workbook	The workbook that medical scribes completed during the virtual medical audit.
Certificate of Confidentiality	Certificate of Confidentiality obtained for the study from the National Institute of Health.
Screening Survey	This survey was completed by each interested facility to confirm study eligibility.
Recruitment Email	This email was sent to jails across the U.S. to recruit them for the study.
Workbook for Scraping News Stories	This workbook was completed by study team members tasked with scraping important jail news stories.