Outside the Perimeter:

Commercial Collection of Crisis Data and the Subpart That Was Never Written

Travis Gilly
ORCID: 0009-0007-2954-6313
Real Safety AI Foundation
Illinois, United States

❧ Working paper, August 2026 (v1.2)

ABSTRACT

American human subjects protection was built around institutions, because the abuses that produced it happened inside institutions. The Common Rule places its review requirement at the point where a covered institution touches a person. This paper argues that the location of that requirement, rather than its substance, is now the operative failure. When the regulated boundary is drawn around the institution, collection can be performed outside it by commercial and quasiinstitutional actors and then delivered inward as material that already exists, at which point the consent question that would have governed the collection is never presented to any review body. The paper names this pattern collection displacement and names the device that carries it the intermediary shield. It traces a documented fortyeight year refusal to write a subpart for adults whose decisional capacity is uncertain or impaired, from the National Commission in 1978 through the National Bioethics Advisory Commission in 1998 to the revised Common Rule of 2018, and shows that even the subpart repeatedly asked for would not reach the present problem, because the collection is performed by parties the rule does not cover. Two instances are examined: the Crisis Text Line and Loris.ai arrangement disclosed in 2022, and a current arrangement in which an unincorporated and then newly incorporated organization solicited conversation records from people in acute psychiatric crisis and supplied them to researchers at multiple universities. The conclusion is that the population most likely to be recruited during a mental health emergency is the one population with no dedicated regulatory protection, and that the collection reaching them now sits outside the perimeter the regulations govern.

Keywords: research ethics; human subjects protection; collection displacement; intermediary shield; institutional review board; decisional capacity; secondary data use; data brokerage; psychiatric crisis; informed consent; digital mental health; Common Rule

I. INTRODUCTION

Every major instrument in American research ethics was written after somebody was hurt, and each was written to reach the place where the hurt occurred. The Nuremberg Code answered physicians acting under a state. The National Research Act of 1974 and the Belmont Report answered a Public Health Service study and the institutional research culture that tolerated it. The Common Rule, codified at 45 CFR Part 46,

answered the question of how a federally supported institution should govern its own investigators. In each case the remedy was placed at the point of contact between an institution and a person.

That placement was correct for the harm it addressed and it has held for fifty years. The argument here is that it no longer reaches the collection that matters most, and that the reason is structural rather than a failure of will by any review board.

The mechanism is simple enough to state in a sentence. A commercial or quasiinstitutional actor, subject to no assurance and no review, solicits sensitive material from people in acute distress. The material is then transferred to university researchers, who receive it as information that already exists. At the receiving end, the analysis turns on whether the information is identifiable. If it is not, the activity falls outside the regulatory definition of human subjects research entirely and no board reviews anything. If it is, one of the secondary research exemptions may apply. Under either route the consent question at the point of collection is never presented to a body with authority to ask it, because the only bodies with that authority sit downstream of the collection.

This paper calls that pattern *collection displacement*: regulation drawn around an institution displaces collection outward rather than eliminating it. It calls the device that makes displacement legible to a review board the *intermediary shield*: a third party whose interposition converts a recruitment problem into a data receipt.

The population where this matters most is the one the regulations have never covered. Subparts exist for pregnant women and neonates, for prisoners, and for children. There is no subpart for adults whose decisional capacity is uncertain or impaired. That absence has been identified and formally objected to since 1978. It survived the 1991 codification, a 2007 public comment solicitation, and the 2018 revision. Section V traces that record. Section VI then shows that even the subpart repeatedly requested would not solve the present problem, because it would bind institutions and the collection has left the institution.

II. Method

The historical material is drawn from published regulatory documents, the reports of the federal advisory bodies that recommended changes, and the secondary literature analyzing the resulting gap. Regulatory text is quoted from the current electronic Code of Federal Regulations rather than from secondary summaries, because the 2018 revision renumbered the definitions and a substantial body of institutional guidance still cites the prior scheme.

Two contemporary instances are examined. The first, Crisis Text Line and Loris.ai, is included because it is fully public, was disclosed in national reporting, produced a published exchange in a peer reviewed journal, and has reached a settled factual record. It functions here as the calibrated case: a reader can check every claim.

The second instance is current and its record is still forming. It is included because it displays the mechanism at a further stage of development, with a collector that solicits directly from a crisis population and multiple receiving institutions in more than one national jurisdiction. Claims about it are sourced in one

of three ways, marked throughout: to a published academic paper, to public corporate and regulatory filings, or to material held by the author. Where a claim rests only on the third category it is stated as such and no inference is drawn beyond the document.

The paper does not attempt a survey of the collection industry. It establishes that the industry exists, using the Duke report on mental health data brokerage, and then argues from the structure rather than from prevalence. Prevalence is named in Section X as the open empirical question this paper does not answer.

III. Before the Perimeter

Two histories converge on the present case and they are usually told separately.

A. The Abuse History

The first is the familiar one. Between 1932 and 1972 the Public Health Service followed several hundred Black men with syphilis at Tuskegee without treating them and without telling them what they had. At Willowbrook State School, children with intellectual disability were deliberately infected with hepatitis, with parental permission obtained under conditions where admission to the institution was scarce. At the Jewish Chronic Disease Hospital in 1963, live cancer cells were injected into chronically ill elderly patients who were not told what was being injected. In 1966 Henry Beecher published a survey of published studies in the New England Journal of Medicine showing that ethical violations of this kind were ordinary rather than exceptional in the mainstream literature.

The structural feature these share is the one that shaped the remedy. Every one of them happened inside an institution, performed by that institution's own investigators, on a population the institution already held. Prisoners, hospitalized patients, and institutionalized children were available because the institution controlled access to them. When Congress and the Department responded, they placed the requirement exactly where the abuse had been: on the institution, at the moment its investigators touched a person.

B. The Longer Strand

The abuse history is usually told as a set of episodes. Read by population rather than by episode, it is one continuous practice.

The people in the Willowbrook hepatitis studies were children with intellectual disability. The patients at the Jewish Chronic Disease Hospital were chronically ill and elderly. Across the middle of the twentieth century, psychiatric hospitals supplied subjects for malaria inoculation, insulin coma, lobotomy, and drug trials, and the supply was reliable because the institution controlled the population and the population's capacity to refuse was treated as already compromised. Before that, asylums furnished bodies for anatomical demonstration and, before that again, the practice reaches back as far as institutional confinement itself.

The common feature is not cruelty. It is availability. A population held by an institution, whose refusals can be characterized as symptoms, is the cheapest research material that has ever existed, and every era has found a reason why using it was defensible at the time.

This matters for Section V. Three populations received dedicated protection in 1991, and each had an organized constituency that could speak in its own name outside the institution holding it. Prisoners had litigation and a civil rights movement. Children had parents. Pregnant women had a political coalition. Adults whose decisional capacity is uncertain or impaired had, and largely still have, no equivalent. The absence of a fourth subpart is not only a drafting omission. It is the same availability, carried forward into the regulatory scheme by the same mechanism that produced it, which is that the affected population is the one least able to object in the forum where objections count.

C. The Panic History

The second history is less often placed alongside the first, and it is the one that predicts the shape of the current case.

Each time a new communication medium reaches wide adoption, a cluster of psychological harm is attributed to it, and a research rush follows in which the affected people become a contested resource. George Beard's neurasthenia, described in 1881, blamed the telegraph, the railway, and the periodical press for an epidemic of nervous exhaustion. The Payne Fund Studies of 1929 to 1933 investigated the effects of motion pictures on children and were criticized for method and for moralizing. Hadley Cantril's 1940 study of the War of the Worlds broadcast established a mass panic in the scholarly record; later work by Pooley and Socolow argued that the panic was substantially a newspaper artifact, which makes Cantril's study the durable evidence for an event the study helped construct. Fredric Wertham's 1954 campaign against comic books drove federal hearings; archival work by Carol Tilley in 2012 found that Wertham had altered and combined the case histories of the children he cited.

Three features recur. The affected population is described by people who did not collect its accounts under any standard. Institutional prestige attaches to whoever reaches the population first. And the research produced during the rush outlives the phenomenon that produced it, becoming the record even where the record was wrong.

What separates the present case from all of them is who does the collecting. In every historical instance the researchers gathered their own material and answered for it, however badly. In the present case the material arrives already gathered.

IV. The Perimeter and Its Three Subparts

The National Research Act of 1974 created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, which produced the Belmont Report in 1979. Belmont set out three principles: respect for persons, beneficence, and justice. Respect for persons carries the requirement of informed consent and, in Belmont's own terms, the requirement of additional protection for persons with diminished autonomy.

The regulatory implementation, codified as Subpart A of 45 CFR Part 46 and adopted across federal departments as the Common Rule, made institutional review the operative device. An institution holds an

assurance, constitutes a board, and the board reviews the protocols of that institution's investigators.

In 1991 the Department added three sets of special protections: Subpart B for pregnant women, human fetuses, and neonates; Subpart C for prisoners; and Subpart D for children. Each of these populations shares the feature that made Tuskegee and Willowbrook possible. Each was a population an institution already held, or could readily reach, and whose capacity to refuse was constrained by that holding.

There is no fourth subpart for adults whose decisional capacity is uncertain or impaired.

V. The Subpart That Was Never Written

The absence is not an oversight discovered later. It was identified at the beginning and has been objected to continuously for fortyeight years.

The National Commission itself, in 1978, published a report recommending that the regulations cover persons described in the language of the period as mentally infirm. The Department did not adopt the recommendation. When the three subparts arrived in 1991, this population was the one that had been asked for and not delivered. Contemporary analysis records that the proposed regulations covering mentally disabled persons were abandoned against strong opposition, leaving research with this population governed only by the general instruction in Subpart A that boards include additional safeguards to protect the rights and welfare of subjects vulnerable to coercion or undue influence.

Through the 1980s and 1990s successive bioethics bodies repeated the request. In December 1998 the National Bioethics Advisory Commission published *Research Involving Persons with Mental Disorders That May Affect Decisionmaking Capacity*, which issued twentyone recommendations addressed to boards, to investigators, and to federal regulators. A Department working group analyzed the report and proposed action. The National Human Research Protections Advisory Committee added recommendations of its own before being disbanded and replaced by the Secretary's Advisory Committee on Human Research Protections, which formed a subcommittee on the inclusion of individuals with impaired decision making and reported in 2009.

In September 2007 the Department published a request for information and comments in the Federal Register asking the public whether the existing protections were adequate for adults with impaired decision making capacity. The solicitation itself is the admission: the Department was asking, twentynine years after the first recommendation, whether the gap identified in 1978 needed filling.

In 2013 Stacey Tovino argued in the Houston Law Review for a new Subpart E governing research involving adults with impaired decision making capacity. The proposal is on the record, in print, with the drafting problem worked through.

The Common Rule was then comprehensively revised, with a final rule in 2017 and general compliance from January 2019. The revision restructured the exemptions, introduced limited review, created broad consent for secondary use, and renumbered the definitions. It did not add the subpart.

The practical consequence has been measured. A cross sectional study of the human subjects policies of ninetyfour top funded American research institutions found that 41.5 percent maintained policies requiring the exclusion of people with uncertain or impaired decision making capacity unless their inclusion is scientifically

justified, while 5.3 percent required inclusion unless exclusion is justified. In the absence of a rule, institutions have converged on avoidance. That is the opposite of protection: the population is neither safeguarded when it is enrolled nor reliably included in the research that concerns it.

VI. Collection Displacement

The gap described above is a gap in what the rule covers. This section describes a gap in where the rule reaches, and the two compound.

A. What the Definition Does

Under 45 CFR 46.102(e)(1), a human subject is a living individual about whom an investigator conducting research either obtains information through intervention or interaction with the individual and then uses, studies, or analyzes that information, or obtains, uses, studies, analyzes, or generates identifiable private information. Identifiable private information, at 46.102(e)(5), is private information for which the subject's identity is or may readily be ascertained by the investigator or is associated with the information.

Two consequences follow directly from the text.

First, the definition attaches to the investigator. It asks what the investigator did, and it does so at the moment of the investigator's contact. The conduct of a noncovered third party who assembled the material before the investigator saw it is not an element of the definition.

Second, where the material reaching the investigator is not identifiable, the second prong is not satisfied and the first was never engaged, because there was no intervention or interaction. The activity is not human subjects research. No board reviews it, no consent standard applies to it, and no record of the collection is created at the institution.

Where the material is identifiable, 46.104(d)(4) exempts secondary research uses of identifiable private information under any of several conditions, including that the information is publicly available, or that the investigator records it so that identity cannot readily be ascertained, does not contact the subjects, and will not reidentify them. This is a sensible provision for the reuse of research and clinical datasets, which is what it was written for.

B. The Intermediary Shield

Place a third party between the person in crisis and the investigator, and the regulatory picture at the receiving institution changes completely.

Without the intermediary, an investigator who solicits conversation records from a person in acute psychiatric distress is interacting with that person, obtaining information, and analyzing it. That is human subjects research on the face of the definition. It requires a protocol, board review, informed consent, and, because of who the subject is, the capacity assessment that no subpart specifies and that Section V shows the regulations have declined for fortyeight years to specify.

With the intermediary, the investigator receives a file. The recruitment was performed by an entity with no assurance. The consent, if any, was obtained by that entity under terms it wrote for itself. The deidentification, if performed, was performed by that entity and is represented to the researchers rather than verified by them. The investigator's own conduct begins at receipt.

This is what the term *intermediary shield* names. It is not a loophole in the sense of a drafting error. It is the predictable operation of a rule whose subject is the institution, applied to a world in which collection is a service that can be bought.

C. The Supply Side Is an Industry

The supply side is not hypothetical and it is not confined to research. In February 2023 the Duke Sanford Cyber Policy Program published a study by Joanne Kim in which the researcher approached thirtyseven data brokers about bulk mental health data. Twentysix responded and eleven were willing and able to sell what was requested. Prices ran from roughly two hundred seventyfive dollars for five thousand aggregated record counts to subscriptions in the range of seventyfive to one hundred thousand dollars covering data on individuals' mental health conditions, with some records carrying names and addresses. The report concluded that the industry lacks best practices for handling this category of data, particularly around privacy and the vetting of buyers. One detail bears on the argument here: every broker asked what the data would be used for, and none of them checked the answer (Enos, 2023).

Purchasing a dataset is an ordinary procurement act inside a research institution. Nothing in the procurement pathway asks how the material at the far end was obtained, because the regulatory question at the institution is about identifiability rather than provenance.

D. What the Literature Already Establishes

The general form of this problem is named in the research ethics literature, and the paper's contribution is narrower than the problem itself.

Friesen and colleagues (2021) trace the institutional review board to a world in which health research was federally funded, performed inside academic institutions, and dependent on data that could only be obtained in clinical settings. Review requirements were built on those features. They then state the consequence directly: research has moved outside clinical and academic contexts, is performed by for-profit device, pharmaceutical, and internet companies that often circumvent the requirement for ethics oversight, and where the data used is de-identified or publicly available the work is treated as exempt regardless of whether risks of re-identification, bias, or downstream reuse are present.

That is collection displacement described from the receiving end. What is added here is the supply end and the population. Friesen and colleagues describe companies that generate data incidentally through their own operations and later make it available. The arrangement examined in Section VIII is different in kind: material is solicited from individuals, for the stated purpose of research, by an entity that exists to solicit it, from people recruited during a psychiatric emergency. Incidental data and solicited data arrive at the same exempt door, and only one of them was gathered by asking a person in crisis to hand it over.

The doctrinal position on the receiving end is settled and uncontroversial among those who study it. Lynch, Bierer and Cohen (2016) summarize it: de-identified data falls outside the threshold definition of human subjects research, secondary research with identified data is exempt where the investigator does not record it identifiably, and waivers of consent for secondary research are routinely granted. Hall and King (2016), reviewing the same proposed revision, concluded that both the existing regime and the proposed replacement were critically flawed. The 2018 rule that emerged kept de-identified data outside the definition entirely.

The health privacy statute does not fill the gap, because it was not built to. Terry (2015) notes that the Privacy Rule reaches protected health information held by covered entities and does not apply to de-identified information at all. White, Blok and Calhoun (2020) put the boundary plainly: health information not held by a covered entity may be used and disclosed without regard to the Privacy Rule. An organization soliciting conversation records from a support community is not a health plan, a clearinghouse, or a provider transmitting claims. It is outside both regimes at once, and this is the ordinary result of two sectoral rules meeting at a party neither describes.

Botkin (2017) identifies what this costs even where every actor complies. Writing about secondary use in learning health systems, he observes that public trust is lost when people discover a practice they were never told about, and that the reaction is not an assessment of risk but a question about disclosure: if this is innocent, why did nobody mention it. He proposes notice and an opt-out. Nothing in the arrangement described in Section VIII offers either.

Finally, the field has already written what good practice looks like for this exact population. Nock and colleagues (2020) report a consensus statement developed through a modified Delphi process with twenty-four experts, including scientists, clinicians, ethicists, legal experts, and people with lived experience, on ethical and safety practice for digital monitoring studies with people at risk of suicide. It addresses inclusion criteria, elements of informed consent, data review during a study, and responding to elevated risk in real time. It is guidance. Nothing requires it, and nothing about the arrangement in Section VIII would have brought it into play.

VII. A Documented Precedent

Crisis Text Line is a text based crisis intervention service. In 2021 a forprofit company, Loris.ai, was launched with a shared founder, and conversation data from the crisis service was used to develop the company's customer service technology. Politico reported the arrangement in January 2022. Within days the service announced that it had ended the data sharing relationship, requested that Loris delete the data it had received, and updated its terms of service.

The published record around it is what makes the case useful.

In 2019 the Journal of Medical Internet Research published a paper by Pisani and colleagues describing the principles of the organization's academic data sharing pilot, authored in part by members of its external data ethics committee. Following the Politico report, a commentary by Tim Reierson raised the conflict

question directly: the organization had a monetary interest in commercial use of the data through its subsidiary, and therefore an interest in the 2019 paper's finding that the crisis conversation data was ethically sourced for research purposes. The journal published Reierson's commentary and a reply from Pisani and colleagues in January 2025, together with an editorial noting that the ethics committee members who authored the 2019 paper were apparently not involved in or consulted about the decision to share data with the commercial subsidiary.

Three features of this case recur in the next one.

The organization published its own ethics assurance and that assurance became the document outside parties relied on. The assurance was produced by people who did not know what the organization was also doing. And the consent instrument at the point of collection was a terms of service acceptance presented to a person in crisis, which a former board chair of the organization later acknowledged in public is not consent.

VIII. The Present Instance

Beginning in 2025, reports of psychological crises associated with extended conversational interaction with large language model systems reached national coverage. As in every case in Section III.B, an affected population formed faster than any institution could reach it, and organizations appeared that solicited accounts directly from that population.

The Human Line Project is one such organization. It solicits conversation records and personal accounts from people who describe crises connected to chatbot use, through intake surfaces that collect names and contact details alongside the records. Its founder describes a support community of more than 450 people across 32 countries, and reports within that population 212 cases of psychosis, more than 109 hospitalizations, and 20 deaths by suicide.

A. The Mechanism, Stated by the Intermediary

At a public event hosted by an AI safety organization in Montreal, recorded and available in full, the founder described the organization's research function in terms that state the mechanism this paper describes. Asked how researchers might connect with the affected population, he answered that the organization does not do research in-house and works only with partners who connect with the community afterward and make their own request of it.

That is the intermediary role described from the inside, without apparent recognition that it is a regulatory position rather than an administrative one. The organization holds the access. Partners arrive after the population has been assembled, and the terms under which it was assembled are not theirs.

In the same session the founder named six universities as research partners: MIT, King's College London, Stanford, McGill, and Yale, among others. All of these associations are therefore sourced to the organization itself, stated in a recorded public session, which is a stronger source than the partner display on its own website and is the source relied on here. The analysis that follows works through the one partnership with a published research output to compare against the organization's account. The provenance question it develops

is not specific to that partnership. On the founder's own description of the mechanism, every named partner arrives after the population has been assembled, on terms the partner did not set, and each therefore faces the same question this paper puts to the one documented instance.

B. Two Accounts of the Same Transfer

The published research record and the founder's account of it do not agree, and the disagreement is the paper's central exhibit.

Moore and colleagues, in a paper on delusional spirals in human and large language model chat logs accepted to the ACM Conference on Fairness, Accountability, and Transparency in 2026, state in their methods that they received chat logs via the organization, describe it as a nonprofit set up as a community for people with lived experience, and state that with individuals' consent the organization removed identifiers from those logs before the research team reviewed them, with no demographic data transferred. Twelve of the study's nineteen participants were sourced this way.

At the recorded session, asked about that study, the founder disclaimed the role entirely. He stated that the study did not go through the organization for the transcripts, that the researchers asked people directly in the community, and that the researchers compensated those people directly. He directed questions about the review board to the university, suggesting that anyone concerned should ask the university how it paid people and how it collected transcripts. He then stated that the paper does not say the organization provided transcripts, only that the researchers had been in contact, and invited the audience to read it.

The paper says otherwise, in the passage quoted above, and any reader can check it.

Later in the same session, answering a different question, the founder referred to the study as one in which the organization analyzes transcripts, described it as being with Stanford University, and reported figures from it. The disclaimer and the claim of participation occur in one recorded conversation.

C. Why the Contradiction Does Not Need Resolving

The paper does not need to determine which account is accurate, because both accounts implicate the same failure from opposite directions.

If the published methods are accurate, the intermediary shield operated exactly as Section VI describes. Consent was obtained by a party under no assurance, de-identification was performed by that party and represented rather than verified, and the receiving institution's regulatory analysis began at receipt. Nobody with authority reviewed the collection, and a subpart for decisional capacity would not have changed that, for the reasons in Section IX.

If the founder's account is accurate, the consequence is the opposite and equally serious. Investigators who approach individuals in a crisis community, enroll them, and pay them are interacting with human subjects under 45 CFR 46.102(e)(1)(i) on the face of the definition. That is not secondary research and no exemption for existing data reaches it. It requires a protocol, review, and informed consent from each person, obtained by the investigators, from a population in acute psychiatric distress, using compensation as an

inducement. That is precisely the enrollment scenario every advisory body since 1978 asked the Department to govern, and the scenario for which no subpart exists.

If the founder's account is accurate, the arrangement also reproduces the historical form described in Section III.A with unusual fidelity. Direct approach to a population assembled by a third party, enrollment of individuals from inside that population, and a modest payment as inducement is the Willowbrook and prison-study shape, minus the physical custody and with the custody function performed by a support community instead. The inducement question is exactly what a review board is meant to weigh for a population whose capacity to refuse may be affected, and it is exactly the weighing for which no subpart supplies a standard.

Either the provenance of this data was never reviewed because an intermediary absorbed it, or it was reviewed under a rule that declines to address the capacity of the people enrolled. The one thing established by both accounts together is that no participant in the arrangement can say with authority how the material was obtained, and that there is no mechanism anywhere in the system for resolving the question. A published methods section and the organization it names give incompatible accounts, and nothing in the regulatory structure requires either to be reconciled with the other.

D. Corroborating Features

Three further features of the recorded session bear on the reliability of the consent representations relied on by receiving institutions.

Asked whether the organization publishes a privacy policy, the founder was unable to confirm that one existed, stated that he does not run the website, and asserted that the organization operates in accordance with Quebec's Law 25. Asked whether a designated person responsible for the protection of personal information is identified publicly, as that statute requires, he stated that the designation exists internally in a job description but did not believe it appeared on the website.

Asked about the organization's corporate structure, and specifically about a for-profit incorporation preceding the nonprofit, he stated that only the nonprofit exists and characterized the earlier incorporation as a transition of a month or two. Public corporate filings, held by the author, record the two entities as separated by a longer interval and as concurrently existing.

The organization's own description of the research findings also departs from the published paper. The founder reported that in 100 percent of analyzed conversations the system presented itself as sentient. The published finding concerns a different subject, and the figure as stated appears to derive from a summary document hosted by the organization rather than from the paper.

None of these establishes bad faith and none is offered as doing so. What they establish is that the consent and de-identification representations on which a receiving institution relied came from an organization that could not, on the record, confirm the existence of its own privacy policy or the public designation its governing statute requires. The regulations ask the receiving institution whether the information is identifiable. They do not ask whether the party that de-identified it has a privacy officer.

E. One Contributor's Account

One contributor states that she asked in writing that her records be kept private, that they were transferred for research regardless, that she then contacted the receiving researcher directly to withdraw them, and that she was refused a copy of her own file. Two withdrawal requests, at two separate nodes of one pipeline, and the material continued at both. That account is held by the author, is reported here with consent limited to this form, and is not published in any identifying form.

It is included because it answers the question the regulatory analysis cannot. The structural argument establishes that no review body sees the collection. This establishes what that absence costs the person at the far end: there is no node in the pipeline at which a withdrawal has to be honored, because no node was ever assigned the duty.

IX. What Should Change

Burris (2008) argued that human subjects protection should be understood as a regulatory space rather than a single instrument, and that the review board is one node among many with the capacity to set standards, monitor conduct, and enforce. His inventory includes funders, accreditation bodies, professional organizations, courts, state law, and journals, each with its own lever. His three framing moves were to recognize the limits of the board as an oversight body, to narrow the range of risks the system is asked to control, and to enroll a wider set of actors.

That framing is what the present problem requires, because the failure is jurisdictional rather than substantive. The material never reaches the body with authority to ask about it. Waiting for a rule that would reach it means waiting on a rulemaking process that has declined the adjacent request for forty-eight years. Four asks follow, ordered by how quickly they can be made and by how little they depend on anyone in Washington.

A. Journals and Conferences

The fastest available lever requires no rulemaking and no new authority. Journals and conference program committees already require a statement of review board approval or a determination of exemption. That requirement should be extended one step, to a provenance statement in the methods: who collected the material, under what instrument, who obtained consent and in what form, and whether the authors verified either or are reporting a representation made to them.

The published methods discussed in Section VIII would satisfy this partly and fail it in the part that matters. They state who collected and who de-identified. They do not state that both facts are representations the authors did not verify, and a reader has to infer it. A required provenance statement makes the inference unnecessary and puts the question in front of the authors while the paper is being written rather than after it is published.

This is the node Burris identifies as holding the sharpest enforcement in ordinary practice: refusal to review or publish. It is also the node most able to act, because a program committee can adopt a submission

requirement in one meeting.

B. Institutional Review Boards

The determination workflow that decides whether an activity is human subjects research currently turns on identifiability. It should also ask provenance: was this material solicited from individuals, by any party, for the purpose of transfer to researchers.

An affirmative answer should not end the analysis or bar the research. It should require the institution to obtain the collection instrument and the consent record before accepting the material. That is a documentation requirement rather than a review requirement, it costs the institution one exchange of files, and it converts a representation into an artifact somebody has read.

C. The Office for Human Research Protections

Guidance is available without rulemaking, and there is precedent for the form. The office has issued guidance on research involving coded private information, which addresses a structurally similar question about when an investigator receiving existing material is engaged in human subjects research.

Parallel guidance should address material solicited by collectors not covered by the rule, state that the exemption for secondary research was written for the reuse of data generated for other purposes, and state that it does not describe material assembled for the purpose of research transfer by a party operating outside any assurance.

D. The Department

Two rulemaking asks, in order of tractability.

First, write the subpart. Tovino (2013) set out the case for a Subpart E governing research involving adults with impaired decision-making capacity and worked through the drafting problem. That argument is on the record and remains correct. Section IX explains why it is insufficient for the arrangement described here; insufficient is not a reason to keep declining it for a fifth decade.

Second, narrow the secondary research exemption at 46.104(d)(4) so that it does not reach information solicited from individuals for the purpose of supply to researchers. The distinction to draw is between material that came into existence for another reason and was later found useful, which is what the exemption was written for and which should keep it, and material assembled by asking people to provide it so that researchers could have it, which is collection wearing the clothing of secondary use.

The line will be contested at its edges and Section XI concedes that. It is nonetheless a line, and it is narrower than the one the 2015 proposed rule attempted and failed to draw across all secondary research at once.

X. Limitations

The prevalence claim is not established here. This paper shows that the mechanism exists, that a market in mental health data exists, and that at least two organizations have operated the mechanism. It does not establish how much research currently rests on displaced collection, and the honest answer is that nobody knows, because the arrangement generates no institutional record by design. Any figure offered would be an estimate presented as a count.

The regulatory analysis in Section VI states the structure of the definitions and their consequence. It does not survey how boards apply them in practice, and board practice varies. It is possible that some institutions ask provenance questions as a matter of local policy even where the regulations do not require it. Evidence of that practice would narrow the claim without defeating it, since local policy is precisely what the absence of a rule leaves to chance.

The strongest argument against the paper is that the intermediary shield is load bearing for legitimate research. Secondary use provisions exist because reusing existing data avoids recontacting subjects, which is itself a protection. A rule that made receiving institutions responsible for the provenance of every dataset would burden ordinary archival and registry research heavily. Any remedy has to separate collection performed for the purpose of supply to researchers from data that came into existence for other reasons, and that line is not obvious.

The account also finds nothing in the ordinary case. A researcher analyzing a public health registry, a census extract, or a curated corpus assembled under an assurance is not doing what this paper describes, and the argument should not be read to reach them.

Finally, the second contemporary instance has a forming record. Several of its elements rest on documents held by the author rather than on published sources, and are marked as such throughout. Independent verification of those elements would strengthen the section and their absence should be weighed by the reader.

XI. Future Work

Two items, each answering a stated limitation.

The prevalence gap is addressable by a bounded audit. Methods sections in the recent literature on digital mental health, self harm, and crisis populations can be coded for whether the data was collected by the research team, received from a named noninstitutional intermediary, or purchased, and for whether the paper states who obtained consent. This is a content analysis over a defined set of venues and years, it needs no new access, and it converts the structural argument into a measured one.

The line between supply collection and incidental data is addressable by a drafting exercise. A rule reaching only material solicited from individuals for the purpose of transfer to researchers, and requiring the receiving institution to obtain the collection instrument and the consent record before accepting it, would leave registry and archival work untouched. Drafting that provision and testing it against the exemption structure at 46.104 is the work that would turn the diagnosis here into a proposal.

XII. Conclusion

The regulations were written against institutions because institutions were where the harm was. Fifty years later the institution is the safest part of the pipeline, because it is the only part anybody watches.

The people supplying the material at the far end of that pipeline are, in the current instance, in psychiatric crisis. They are the one population every advisory body since 1978 has asked the Department to protect specifically, and the one population for which no subpart was ever written. They are now being reached by parties that no subpart would govern, through instruments the parties draft themselves, and their accounts arrive at universities as files with the provenance question already closed.

The remedy does not wait on a rulemaking. A journal can require a provenance statement in the methods section, and a program committee can adopt that requirement in a single meeting. An institution can ask, before it accepts a dataset, who solicited this and what were they told. Those two questions would have reached every case in this paper.

None of this requires anyone to have acted in bad faith, which is what makes it durable. A researcher who receives a deidentified dataset and asks no further questions is complying with the rule as written. A board that never sees the transaction cannot review it. An organization that solicits from a crisis community and writes its own consent terms is doing something no federal rule prohibits. The harm is distributed across a set of individually defensible steps, and the person whose worst hours are in the file is the only party to the arrangement who is not represented anywhere in it.

ACKNOWLEDGMENTS

The author discloses the use of speech-to-text software (Wispr Flow) and Anthropic's Claude as assistive technology for a diagnosed neurodevelopmental condition. All intellectual contributions, research questions, theoretical arguments, methodological decisions, and conclusions are solely the author's own.

Correspondence concerning this article should be addressed to Travis Gilly, Real Safety AI Foundation, Illinois, United States. Email: t.gilly@ai-literacy-labs.org.

References

Abdillahi, I. (2024). Algorithmic surveillance in the era of the mental health appsphere. *American Journal of Community Psychology*, 73(1–2), 17–26. https://doi.org/10.1002/ajcp.12743

Beard, G. M. (1881). *American nervousness: Its causes and consequences*. G. P. Putnam's Sons.

Beecher, H. K. (1966). Ethical problems and clinical research. *New England Journal of Medicine*, 274(24), 1354–1360.

Botkin, J. R. (2017). Transparency and choice in learning healthcare systems. *Learning Health Systems*, 2(1). https://doi.org/10.1002/lrh2.10049

Burris, S. (2008). Regulatory innovation in the governance of human subjects research: A cautionary tale and some modest proposals. *Regulation & Governance*, 2(1), 65–84. https://doi.org/10.1111/j.1748-5991.2007.00025.x

Cantril, H. (1940). *The invasion from Mars: A study in the psychology of panic*. Princeton University Press.

Crisis Text Line. (2022, January 31). *An update on data privacy, our community and our service*. https://www.crisistextline.org/blog/2022/01/31/an-update-on-data-privacy-our-community-and-our-service/

Enos, G. (2023). Duke public policy report suggests federal law restricting data brokers. *Mental Health Weekly*, 33(10), 1–7. https://doi.org/10.1002/mhw.33562

Eysenbach, G. (2025). Crisis Text Line and Loris.ai controversy highlights the complexity of informed consent on the internet and data-sharing ethics for machine learning and research. *Journal of Medical Internet Research*, 27, e67878.

Kim, J. (2023). *Data brokers and the sale of Americans' mental health data*. Duke University Sanford School of Public Policy, Cyber Policy Program.

Friesen, P., Douglas-Jones, R., Marks, M., Pierce, R., Fletcher, K., Mishra, A., Lorimer, J., Véliz, C., Hallowell, N., Graham, M., Chan, M. S., Davies, H., & Sallamuddin, T. (2021). Governing AI-driven health research: Are IRBs up to the task? *Ethics & Human Research*, 43(2), 35–42. https://doi.org/10.1002/eahr.500085

Hall, M. A., & King, N. M. (2016). Conscience, courage, and "consent." *Hastings Center Report*, 46(2), 30–32. https://doi.org/10.1002/hast.546

Horizon Omega. (2026). *Etienne Brisson (The Human Line Project) at the Montreal AI Safety, Ethics and Governance Event Series* [Audio recording of public livestream]. Internet Archive. https://archive.org/details/brisson-human-line-project-montreal-ai-safety-2026_202608

Livingstone, S., Orben, A., & Odgers, C. (2022). Debate: Should academics collaborate with digital companies to improve young people's mental health? *Child and Adolescent Mental Health*, 28(1), 150–152. https://doi.org/10.1111/camh.12630

Lynch, H. F., Bierer, B. E., & Cohen, I. G. (2016). Confronting biospecimen exceptionalism in proposed revisions to the Common Rule. *Hastings Center Report*, 46(1), 4–5. https://doi.org/10.1002/hast.528

Moore, J., Haber, N., et al. (2026). Characterizing delusional spirals through human-LLM chat logs. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*.

National Bioethics Advisory Commission. (1998). *Research involving persons with mental disorders that may affect decisionmaking capacity*.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. Department of Health, Education, and Welfare.

Nock, M. K., Kleiman, E. M., Abraham, M., Bentley, K. H., Brent, D. A., Buonopane, R. J., et al. (2020). Consensus statement on ethical and safety practices for conducting digital monitoring studies with people at risk of suicide and related behaviors. *Psychiatric Research and Clinical Practice*, 3(2), 57–66. https://doi.org/10.1176/appi.prcp.20200029

Pisani, A. R., Gallo, C., Gould, M. S., Kanuri, N., Marcotte, J. E., Pascal, B., Rousseau, D., Ranney, M. L., Filbin, B., & Turner, S. (2025). Authors' reply: Commentary on protecting user privacy and rights in academic data-sharing partnerships. *Journal of Medical Internet Research*, 27, e59734.

Pooley, J., & Socolow, M. (2013). The myth of the War of the Worlds panic. *Slate*.

Terry, N. (2015). Developments in genetic and epigenetic data protection in behavioral and mental health spaces. *Behavioral Sciences & the Law*, 33(5), 653–661. https://doi.org/10.1002/bsl.2203

Tilley, C. L. (2012). Seducing the innocent: Fredric Wertham and the falsifications that helped condemn comics. *Information and Culture*, 47(4), 383–413.

Tovino, S. A. (2013). A "common" proposal. *Houston Law Review*, 50, 787.

United States Department of Health and Human Services. (2007). Request for information and comments on research that involves adult individuals with impaired decision-making capacity. *Federal Register*, 72, 50966.

United States Department of Health and Human Services. (2018). Protection of human subjects, 45 C.F.R. pt. 46.

Wertham, F. (1954). *Seduction of the innocent*. Rinehart.

White, T., Blok, E., & Calhoun, V. D. (2020). Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. *Human Brain Mapping*, 43(1), 278–291. https://doi.org/10.1002/hbm.25120