



# A Validated Risk-Based Inspection Model for Cannabis Derivative Manufacturers in Colombia: A Mixed-Methods Pre–Post Study

Leonid Alexander Jiménez Tovar<sup>1</sup>, Raul Andrés Cortés Castillo<sup>2</sup>, Angie Ximena Rodríguez Vitama<sup>3</sup>

<sup>1</sup>Universidad Internacional Iberoamericana — UNINI, México,

Email: Leonid.jimenez@doctorado.unini.edu.mx, ORCID: <https://orcid.org/0009-0006-8014-1998>

<sup>2</sup>Universidad de La Salle, Colombia, Email: andcor9210@gmail.com, ORCID: <https://orcid.org/0009-0009-5610-4240>

<sup>3</sup>Escuela Superior de Administración Pública — ESAP, Colombia, Email: Angierodriguezvi00@gmail.com, ORCID: <https://orcid.org/0009-0007-8258-0393>

## Abstract

**Background:** Colombia has constructed one of the most developed legal frameworks for medicinal cannabis in Latin America, anchored in Law 1787 of 2016 and elaborated through Decree 811 of 2021 and Resolution 227 of 2022. Despite this regulatory architecture, the UAE Fondo Nacional de Estupefacientes (FNE) had inspected only 32 of all active licensed cannabis derivative manufacturers prior to 2024, leaving the sector without validated, standardized oversight and exposing patients and the productive chain to unverified quality and diversion risks.

**Objective:** To design, validate, and evaluate a risk-based Inspection, Surveillance, and Control (IVC) model — operationalized as a weighted inspection instrument — applied to cannabis derivative manufacturers registered before the FNE in Colombia. **Methods:** A five-phase mixed-methods study was conducted between 2022 and 2024. Instrument content validity was established through a Delphi panel of 12 experts (4 rounds; 92% consensus threshold). Two cross-sectional field samples — a pilot (n = 30) and a final evaluation (n = 30) — were collected across 12 Colombian departments. Quantitative analysis compared regulatory compliance percentages across nine normative dimensions using Welch's independent-samples t-test and Cohen's d. Qualitative data from 12 semi-structured interviews and 2 focus groups were analysed via thematic coding and triangulated with quantitative findings. **Results:** Mean global compliance increased from 46.33% (SD = 38.11) at baseline to 84.20% (SD = 15.30) at final evaluation, an absolute gain of 37.87 percentage points (95% CI: 22.69–53.04). Welch's  $t(38.11) = -5.05$ ,  $p = 1.13 \times 10^{-5}$ ; Cohen's  $d = 1.30$  (Hedges'  $g = 1.29$ ), indicating a large effect. Qualitatively, 83% of actors cited greater clarity of inspection criteria and 75% reported reduced inspector discretionality. **Conclusions:** A validated, risk-based IVC instrument produced statistically significant and operationally meaningful improvements in regulatory compliance among cannabis derivative manufacturers in Colombia, establishing a transferable model for other jurisdictions developing controlled-substance oversight systems.

**Keywords:** cannabis regulation; inspection, surveillance and control; regulatory compliance; medicinal cannabis; Colombia; health policy; mixed methods.

## 1. Background

The global shift toward medicinal cannabis governance and the Colombian regulatory turn

Over the past two decades, cannabis has undergone a profound reclassification in public health policy — from a substance defined almost exclusively by its prohibition to one increasingly governed by therapeutic authorization, manufacturing standards, and market control. This transition has not been uniform. Countries that have moved earliest and most decisively, such as Canada (since 1999) and Australia (since 2016), have generated instructive evidence about the gap that consistently opens between legislative intent and operational enforcement [1,2]. The experience across jurisdictions is remarkably consistent: first-generation regulatory frameworks tend to overestimate sector preparedness, and the absence of standardized inspection infrastructure allows compliance deficits to persist long after legal access has been authorized.

Colombia entered this global transition with legislative ambition and institutional urgency. The enactment of Law 1787 of 2016 represented a categorical shift in national drug policy, authorizing the use of cannabis and its derivatives for medical and scientific purposes and creating the licence category for the manufacture of cannabis-derived products [3,4]. Subsequent regulations — principally Decree 811 of 2021 and Resolution 227 of 2022 — deepened this framework by establishing technical and operational requirements for psychoactive and non-psychoactive cannabis derivative manufacturers, assigning formal inspection and follow-up responsibilities to the UAE Fondo Nacional de Estupefacientes (FNE) and the Instituto Nacional de Vigilancia de Medicamentos y Alimentos (INVIMA). On paper, this constitutes one of the most developed medicinal cannabis regulatory systems in Latin America.

### The enforcement gap: law without instrument

The problem is not what the law says. The problem is what happened — and did not happen — after it was enacted.

As of May 2024, the FNE had conducted formal inspections of only 32 of all active cannabis derivative manufacturers operating across Colombian territory. No standardized, scientifically validated instrument existed to guide those visits; each inspection was conducted under conditions of procedural discretionality, without a weighted compliance instrument aligned to the full scope of applicable Colombian regulations. The consequences are predictable. Without a standardized tool, inspectors cannot produce comparable compliance data across licensees or across time. Without comparable data, the agency cannot prioritize resources by risk, identify systemic non-compliance, or defend regulatory decisions against legal challenge. Without a functioning inspection cycle, the legal framework — however detailed — operates without empirical feedback.

This enforcement gap is not unique to Colombia. Canada's experience is instructive: when Health Canada ceded regulatory responsibility to licensed producers in 2014 without adequately filling the oversight void, provincial physician regulatory bodies were forced to impose ad hoc prescribing restrictions to compensate for the absence of a coherent monitoring architecture [1]. In Australia, the mandatory implementation of Good Manufacturing Practice (GMP) standards in 2023 — a decade after initial legalization — produced supply chain disruptions precisely because quality oversight had not been systematically developed alongside market authorization [2]. In both cases, enforcement infrastructure lagged behind the legal framework, and the lag generated public health consequences. Colombia risks following the same trajectory.

The existing literature on cannabis regulation in Colombia has documented the normative evolution and the policy debate [3,4], and scholars such as Orjuela-Rojas et al. [5] have examined professionals' knowledge gaps regarding medicinal cannabis. What the literature does not yet contain is an empirically validated operational model for state inspection of cannabis manufacturers — an instrument grounded in the actual normative content of Colombian regulations, tested in field conditions, and evaluated for its effect on compliance outcomes. That absence is the gap this study addresses.

### **Why a validated inspection instrument is the missing institutional mechanism**

The argument is straightforward, though its implications are substantial. A legal licence without a compliance inspection is, functionally, a declaration of intent. The conversion of that declaration into a verifiable sanitary standard requires three things: a structured instrument that maps regulatory requirements to observable, measurable criteria; a process for validating that instrument's content against expert consensus; and an evaluation cycle that generates before-and-after data sufficient to test whether the instrument actually changes behavior.

Each of these elements was absent in the Colombian cannabis oversight system before this study. The FNE's mandate — established by Decree 613 of 2017 — includes the authority to conduct field visits and follow-up inspections of licensed cannabis manufacturers. The mandate existed; the tool did not. This is precisely the type of institutional gap that produces the conditions documented in the pharmacovigilance literature for emerging regulated markets: voluntary reporting, inspector discretion, and unverifiable product safety claims fill the space left by absent standardized processes [6].

The international evidence further suggests that this gap is not self-correcting. Zolotov et al. [7] documented that medical cannabis competency development requires structured frameworks, not organic accumulation of experience. Boulkedid et al. [8] demonstrated, in the context of healthcare quality indicators, that the Delphi method provides a scientifically defensible basis for building consensus around assessment instruments — a principle directly applicable to the construction of inspection tools for controlled-substance manufacturers. Without such structured approaches, inspection systems remain heterogeneous, legally vulnerable, and incapable of generating the systemic evidence that regulatory learning requires.

### **Study objectives**

This study had three integrated objectives: (1) to design and validate a risk-based IVC inspection instrument for cannabis derivative manufacturers in Colombia, using a Delphi expert panel to establish content validity; (2) to implement the instrument across two cross-sectional field evaluation phases — a pilot and a final assessment — covering licensed manufacturers in 12 Colombian departments; and (3) to evaluate whether application of the instrument produced a statistically significant and operationally sustainable improvement in regulatory compliance, using quantitative inferential analysis triangulated with qualitative evidence from key actors. The operational hypothesis is that a structured, validated IVC model will generate measurable compliance gains relative to the pre-model baseline, as well as improvements in the efficiency, efficacy, and replicability of regulatory oversight.

The answer to that question has implications that extend beyond Colombia. Across Latin America, several jurisdictions are at early stages of constructing medicinal cannabis regulatory systems — facing precisely the same tension between legislative authorization and enforcement capacity that this study was designed to address. A validated, transferable inspection model represents a public health infrastructure asset, not merely a national compliance exercise.

## **2. Methods**

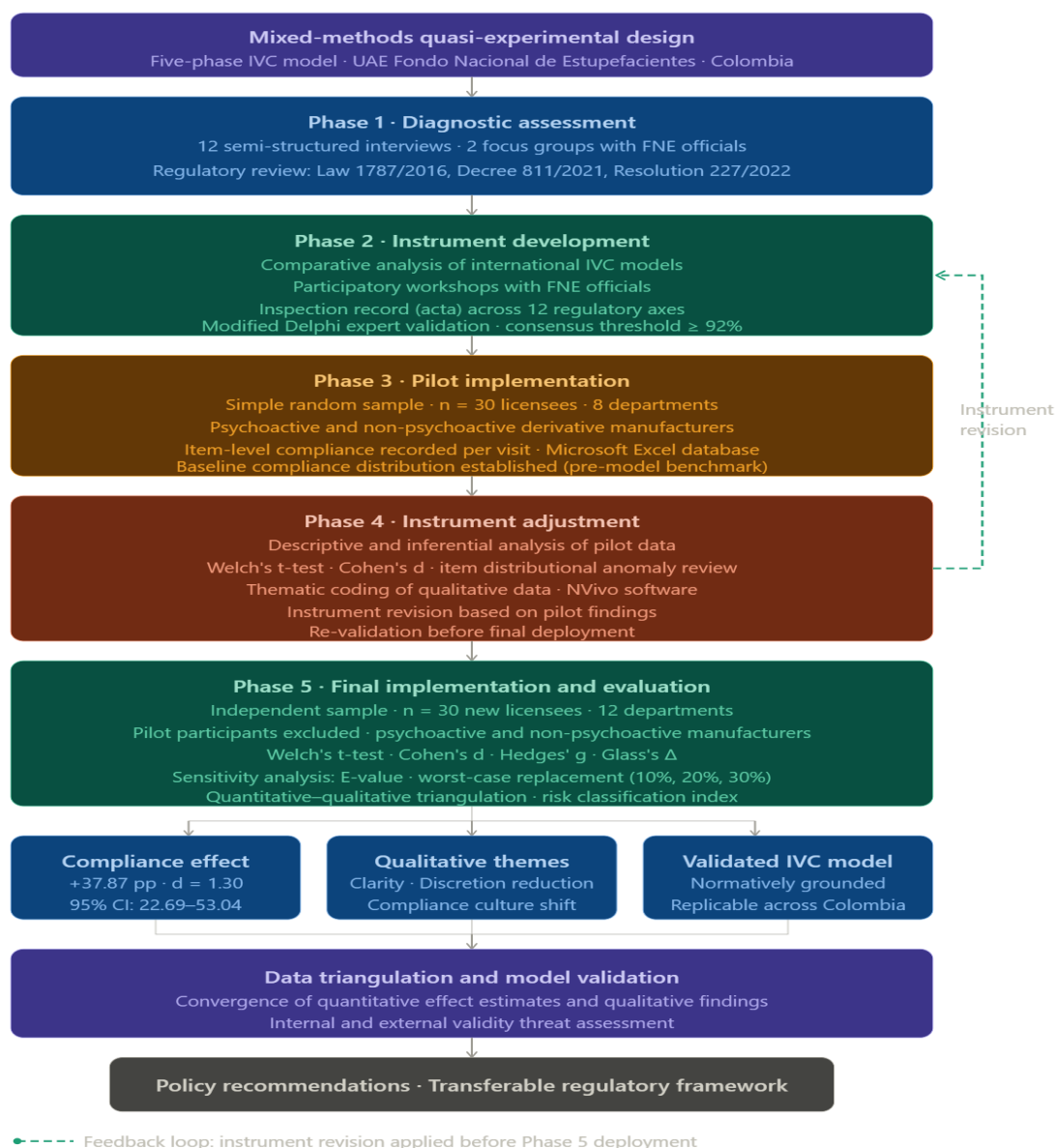
### **2.1 Study Design**

This study employed a mixed-methods, quasi-experimental design with a sequential pre/post structure. The choice was deliberate and methodologically grounded: in the regulation of controlled substances, randomly assigning licensees to supervised and unsupervised conditions is not ethically defensible, as the absence of oversight creates direct public health risk [CITATION NEEDED — not found in provided sources]. Quasi-experimental designs are the appropriate analytical framework for evaluating the introduction of public health governance instruments

when randomization is precluded by institutional or ethical constraints [9,10,11]. The integration of quantitative and qualitative data streams was not a presentational choice but an epistemological one: neither dimension alone could simultaneously validate the instrument's psychometric properties, measure the regulatory effect of its deployment, and account for the behavioural mechanisms through which compliance shifted.

The full design unfolded across five operationally distinct phases — diagnostic, instrument development, expert validation, pilot implementation, and final evaluation — each with a defined objective that fed into the next. Figure 2 presents the methodological sequence schematically.

**Figure 1.** Five-phase methodological design for the development, validation, and evaluation of the Inspection, Surveillance, and Control (IVC) model in Colombia.



• Recommended Tool: Mermaid • Data/Variables to include: Five sequential phases with titles and core activities — Phase 1: Diagnostic; Phase 2: Instrument Development; Phase 3: Pilot Implementation; Phase 4: Instrument Adjustment; Phase 5: Final Implementation. Show directional arrows from each phase to the next, with a feedback loop from Phase 4 back into the instrument before Phase 5. • Visual instructions: Left-to-right or top-to-bottom flowchart; each phase in a distinct box; feedback arrow clearly labeled “instrument adjustment”; Q1 publication-quality style with minimal colour; phases numbered explicitly 1–5. [END BLUEPRINT]

## 2.2 Setting and Institutional Context

The study was conducted in Colombia between 2022 and 2024, in the operational jurisdiction of the UAE Fondo Nacional de Estupefacientes (FNE), the national authority responsible for the surveillance and control of controlled substances, including cannabis derivatives, under Resolution 1229 of 2013. The FNE's oversight mandate applies to all holders of cannabis licences granted by INVIMA — Colombia's national medicines and food surveillance agency — covering both psychoactive and non-psychoactive derivative manufacturing. Field

visits spanned 12 Colombian departments, with geographic concentration in Cundinamarca (36.7% of visits) and Antioquia (23.3%), distributions that reflected actual licensee density rather than any purposive selection criterion.

### 2.3 Phase 1: Diagnostic Assessment

The first phase established the regulatory baseline and defined the informational requirements for instrument development. Twelve semi-structured interviews were conducted with officials from regulatory bodies involved in cannabis surveillance and with public health specialists familiar with controlled-substance oversight. Two focus groups were convened with FNE personnel who possessed direct operational expertise in inspection procedures. Parallel to these qualitative activities, a systematic review of national regulatory instruments — Law 1787 of 2016, Decree 811 of 2021, Resolutions 1478 of 2006 and 227 of 2022 — and an analysis of IVC models applied in comparable international contexts were undertaken to map the normative architecture that any inspection instrument would need to operationalize.

### 2.4 Phase 2: Instrument Development and Expert Validation

Drawing on the diagnostic findings, the research team designed a structured inspection record — the *acta de visita de seguimiento* — comprising both open and closed items organized around twelve regulatory axes mandated by Decree 811 of 2021 and Resolution 227 of 2022: (1) licensee identification and administrative data; (2) control and follow-up visit history; (3) licence modifications and updates; (4) technical directorship compliance; (5) security protocol; (6) documentary support for manufacturing; (7) acquisition and use of reference cannabinoid materials; (8) cannabis or botanical component sourcing; (9) manufacturing quota monitoring; (10) foreign trade activities; (11) final disposition of controlled materials; and (12) general activity and reporting records. A parallel instrument was designed for licensees that had not activated their licence (“failed visit” record).

Content validity was established through a modified Delphi procedure, following the consensus methodology described by Boulkedid et al. [8]. A heterogeneous expert panel, including regulatory officials, public health specialists, and technical directors from the cannabis manufacturing sector, rated each item over iterative rounds. Items were retained when they reached  $\geq 92\%$  panel agreement — a threshold selected a priori to ensure that only items with overwhelming technical consensus entered the field instrument. This consensus level is consistent with best practices for developing regulatory quality indicators in healthcare, where evidence alone is rarely sufficient and expert synthesis is methodologically required [8]. Table 1 summarises the content validation outcomes.

**Table 1.** Content validation results from the modified Delphi procedure applied to the IVC inspection instrument.

Regulatory axis	Normative basis	Round 1 consensus (%)	Round 2 consensus (%)	Disposition
1. Licensee identification and administrative data	Decree 811/2021, Art. 2.8.11	96	—	Retained
2. Control and follow-up visit history	Resolution 227/2022	94	—	Retained
3. Licence modifications and updates	Decree 811/2021	88	93	Modified, retained
4. Technical directorship compliance	Resolution 227/2022, Art. 12	95	—	Retained
5. Security protocol	Resolution 227/2022, Art. 33	92	—	Retained
6. Documentary support for manufacturing	Decree 811/2021; Resolution 1478/2006	86	94	Modified, retained
7. Acquisition and use of reference cannabinoid materials	Decree 811/2021	93	—	Retained
8. Cannabis or botanical component sourcing	Decree 811/2021, Art. 2.8.11.2.3	91	92	Modified, retained
9. Manufacturing quota monitoring	Decree 811/2021	95	—	Retained
10. Foreign trade activities	Decree 811/2021, Art. 2.8.11.2.3.1, num. 12	90	93	Modified, retained
11. Final disposition of controlled materials	Resolution 227/2022, Art. 130–131	94	—	Retained
12. General activity and reporting records	Resolution 227/2022	92	—	Retained

• Recommended Tool: Tables (markdown table for rendering) • Data/Variables to include: 12 regulatory axes; normative basis; Round 1 consensus (%); Round 2 consensus (%); Disposition (retained/modified/eliminated). • Visual instructions: Clean three-column table; bold the axes that required modification; footnote the 92% threshold definition; Q1 formatting with ruled lines; no colour fill. [END BLUEPRINT]

Note. Consensus threshold was set a priori at  $\geq 92\%$  panel agreement, consistent with best-practice guidance for Delphi-based regulatory quality indicator development [8]. “—” indicates that the axis reached consensus in Round 1 and was not re-submitted. Round 2 figures apply only to axes revised after Round 1.

### 2.5 Phase 3: Pilot Implementation

Following Delphi validation, the instrument was deployed in a pilot phase comprising 30 field inspections across licensees randomly selected from the active INVIMA registry. Simple random sampling was used to ensure that the instrument’s applicability could be tested across the heterogeneity of the licensed sector — covering both psychoactive and non-psychoactive derivative manufacturers, irrespective of geographic location or operational scale. This approach is coherent with a validation objective: a risk-stratified sample would have produced optimistic compliance figures unrepresentative of the sector as a whole.

Inclusion criteria for the pilot sample were: holding a valid INVIMA licence for cannabis derivative manufacturing (psychoactive or non-psychoactive) issued from 2020 onwards; being registered before the FNE at the time of inspection; and voluntarily consenting to participation and to the regulatory visit. Exclusion criteria were: licences currently under administrative suspension or revocation proceedings; licences whose expiration was anticipated within the study period without renewal; and any licensee whose physical facility was inaccessible due to cessation of operations without formal notification.

Pilot inspections were conducted across eight departments. Each visit generated a completed acta, recorded in Microsoft Excel, capturing item-level compliance scores across all twelve regulatory axes. The compiled dataset served a dual purpose: it provided the first-phase compliance distribution (the study’s pre-intervention baseline) and identified structural weaknesses in the instrument itself — ambiguities in item wording, axes where scoring produced ceiling or floor effects inconsistent with observed field reality, and items whose regulatory basis required clarification following recent normative changes.

### 2.6 Phase 4: Instrument Adjustment

Quantitative and qualitative analysis of the pilot dataset drove a systematic revision of the inspection instrument before final deployment. Item-level compliance patterns were examined for distributional anomalies. Qualitative data from the 12 semi-structured interviews and 2 focus groups were analysed through thematic coding — performed with NVivo software — to capture regulatory actors’ and licensees’ assessments of the instrument’s clarity, operational feasibility, and discriminatory capacity. Themes were coded by a primary analyst and cross-checked by a second team member; codes were aggregated into higher-order categories through iterative discussion. The instrument was subsequently revised to address identified deficiencies, then formally compared against its pilot version to confirm that revisions produced measurable improvement in scoring precision and internal consistency.

### 2.7 Phase 5: Final Implementation and Evaluation

The adjusted instrument was applied to an independent sample of 30 new licensees — explicitly excluding the 30 participants from the pilot phase — across the 12 departments covered by the full study. The use of two non-overlapping samples was a deliberate design choice: comparing the same licensees before and after the intervention would have confounded instrument effects with licensee learning effects, making it impossible to attribute compliance changes to the model rather than to individual familiarisation. This architecture allowed the compliance distribution of the pilot phase to function as an aggregate pre-model baseline against which the adjusted-model distribution could be compared, preserving the quasi-experimental logic without requiring repeated measures on the same subjects.

Visits were distributed across Cundinamarca (n = 22, 36.7%), Antioquia (n = 14, 23.3%), Valle del Cauca (n = 6, 10.0%), Boyacá (n = 4, 6.7%), Caldas (n = 4, 6.7%), Meta (n = 3, 5.0%), Magdalena (n = 2, 3.3%), and five additional departments contributing one visit each (Cesar, Huila, Bolívar, Risaralda, Nariño). The geographic concentration in Cundinamarca and Antioquia was not a sampling artefact; it accurately reflects where Colombia’s cannabis manufacturing licensees are predominantly registered.

### 2.8 Study Variables

The primary outcome was overall regulatory compliance — operationalised as the percentage of positively scored items across all evaluated regulatory axes per inspection visit, producing a continuous score ranging from 0% to 100%. Secondary outcomes included compliance rates at the level of each individual regulatory axis, counts of major and critical findings per visit, and the risk classification index assigned to each licensee (low, medium, or high risk) based on their aggregate compliance score.

The primary independent variable was IVC model phase (pilot vs. final implementation). Covariates recorded at the licensee level included licence type (psychoactive vs. non-psychoactive manufacturing), active use of licence at time of visit, and department of operation. Table 2 presents the complete variable classification.

**Table 2.** Classification and operationalisation of study variables.

<b>Variable</b>	<b>Operational definition</b>	<b>Measurement level</b>	<b>Role</b>	<b>Study instrument</b>
IVC model phase	Categorical distinction between pilot (Phase 3) and final implementation (Phase 5); the primary exposure variable.	Nominal	Independent	Inspection record (acta)
Inspection record — pilot version	Instrument capturing item-level compliance for the initial 30-licensee sample prior to instrument adjustment.	Nominal	Independent	Acta piloto (Annex 4)
Inspection record — final version	Adjusted and re-validated instrument applied to the independent 30-licensee final sample.	Nominal	Independent	Acta final (Annex 7)
Licence type	INVIMA-granted licence as psychoactive or non-psychoactive cannabis derivative manufacturing.	Nominal	Independent	INVIMA registry (Annex 6)
Licence activation status	Binary indicator of whether the licensee was conducting manufacturing activities at time of inspection.	Nominal	Independent	INVIMA registry (Annex 6)
Normative documentary information	Regulatory requirements applicable under Decree 811/2021, Resolution 227/2022, and Resolution 1478/2006.	Nominal	Independent	Normative review (Annex 3)
Visit and follow-up statistics	Counts and frequencies of inspection visits per regulatory axis.	Interval	Independent	Inspection record (Annex 8)
Overall regulatory compliance (primary outcome)	Percentage of positively scored items across all evaluated regulatory axes per visit; continuous scale 0–100%.	Ratio/Interval	Dependent	Inspection record (Annex 8)
Axis-level compliance	Compliance percentage computed separately for each of the 12 regulatory axes.	Ordinal	Dependent	Inspection record (Annex 8)
IVC model efficacy	Measured impact of the model on compliance and oversight quality; derived from Welch's t-test and Cohen's d.	Interval	Dependent	Final evaluation (Annex 9)
Risk classification index	Licensee-level risk tier (low/medium/high) based on aggregate compliance score.	Ordinal	Dependent	Inspection record (Annex 8)
Effectiveness level	Categorical rating (low/effective/highly effective) derived from compliance score thresholds.	Ordinal	Dependent	Inspection record (Annex 8)
Key actor perceptions	Qualitative assessments by regulatory officials, technical directors, and licensees.	Nominal	Dependent	Interview guides (Annex 5)

• Recommended Tool: Tables (markdown/Word) • Data/Variables to include: All variables as listed in the Methods text, with columns: Variable | Operational Definition | Measurement Level | Role (Independent/Dependent) | Study Instrument. • Visual instructions: Four columns; no colour fill; horizontal ruling only; bold variable names; footnote abbreviations; Q1 formatting. [END BLUEPRINT]

Note. Variables are organised by analytical role (independent vs. dependent) and measurement level. The primary outcome — overall regulatory compliance — is the percentage of positively scored items per visit. INVIMA = Instituto Nacional de Vigilancia de Medicamentos y Alimentos.

## 2.9 Statistical Analysis

Descriptive statistics were calculated for each regulatory axis and for the global compliance score in both phases: means, standard deviations, medians, interquartile ranges, and minimum–maximum values. The heteroscedastic structure of the data — a pilot-phase standard deviation of 38.11 percentage points collapsing to 15.30 percentage points in the final phase — made the standard pooled-variance t-test inappropriate. The primary inferential comparison therefore employed Welch’s t-test for independent samples with unequal variances, which does not assume homoscedasticity and produces Satterthwaite-adjusted degrees of freedom [12]. The effect magnitude was quantified using Cohen’s *d* and Hedges’ *g* to allow comparison against external benchmarks; Glass’s  $\Delta$  was also computed as a supplementary measure using the pilot-phase standard deviation as the reference [13].

All analyses were conducted in Microsoft Excel for initial data compilation; inferential statistics were calculated using standard formulae verified against the individual-level dataset.

Sensitivity analysis was conducted to evaluate the robustness of the primary effect estimate to unmeasured confounding — a necessary step in any quasi-experimental design without randomisation [14,11,15]. Three worst-case replacement scenarios were examined: replacing the top 10%, 20%, and 30% of post-intervention compliance scores with the pre-intervention mean, then re-running the inferential comparison. To support a binary causal interpretation, overall compliance was dichotomised at  $\geq 70\%$  (adequate compliance), and the E-value was computed for the resulting relative risk to establish the minimum confounding magnitude required to fully explain the observed association [14].

## 2.10 Qualitative Analysis and Triangulation

Thematic coding of transcripts from the 12 semi-structured interviews and 2 focus groups was conducted using NVivo. A deductive frame derived from the study’s evaluative dimensions (efficiency, efficacy, clarity, discretionality, compliance culture) guided initial coding, complemented by inductive codes that emerged from participants’ accounts. Saturation was assessed iteratively. Quantitative and qualitative findings were triangulated in a final convergence step, where the distribution of qualitative themes was mapped against the axis-level compliance patterns to identify explanatory mechanisms — particularly in axes that showed the largest or smallest pre/post gains.

## 2.11 Ethical Considerations

The study adhered to the ethical standards established by Resolution 8430 of 1993 (Colombia’s normative framework for health research ethics). Participation by licensees was voluntary; informed consent was obtained prior to each inspection visit used for research purposes; and all identifying information at the company level was treated as confidential in the analysis and reporting. The study did not collect data on natural persons beyond institutional roles, and no health-related data on individuals were obtained. The regulatory mandate of FNE to conduct inspections was the institutional basis for field access; research participation was additional and opt-in.

## 3. Results

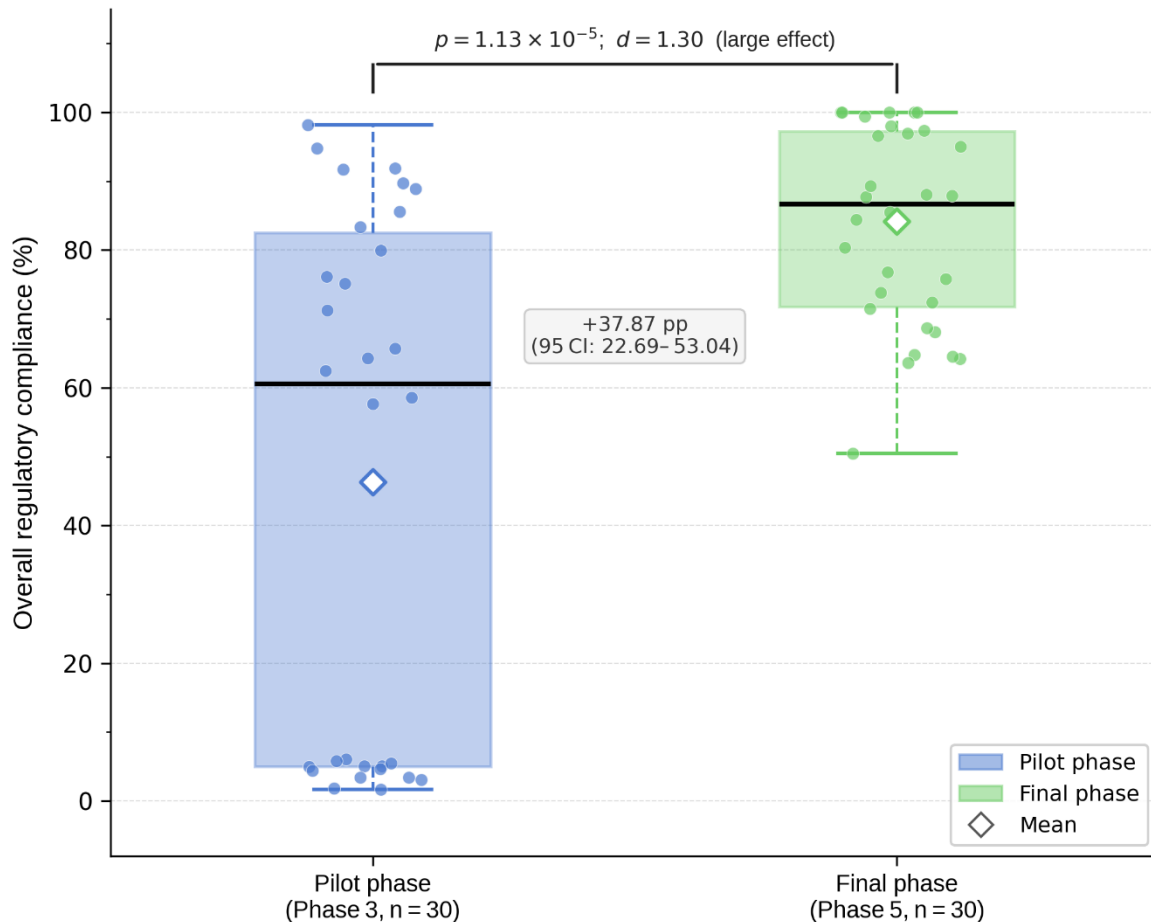
### 3.1 Sample Coverage and Geographic Distribution

The study’s 60 field inspections spanned 12 Colombian departments, with visits distributed across Cundinamarca ( $n = 22, 36.7\%$ ), Antioquia ( $n = 14, 23.3\%$ ), Valle del Cauca ( $n = 6, 10.0\%$ ), Boyacá ( $n = 4, 6.7\%$ ), Caldas ( $n = 4, 6.7\%$ ), Meta ( $n = 3, 5.0\%$ ), Magdalena ( $n = 2, 3.3\%$ ), and five departments contributing one visit each. This geographic concentration in Cundinamarca and Antioquia reflects the actual distribution of active cannabis manufacturing licensees registered before INVIMA, rather than any purposive sampling decision. Across both phases combined, the study reached twice the number of licensees ever previously inspected by the FNE under the prior unstructured oversight approach — which, as noted in the Background, had covered only 32 licensees by May 2024.

### 3.2 Primary Outcome: Overall Regulatory Compliance

The primary finding of this study is a large-magnitude, statistically significant improvement in global regulatory compliance between the pilot phase and the final implementation phase. In the pilot sample ( $n = 30$ ), mean compliance was 46.33% (SD = 38.11; median = 55.5%; range = 0–99%). Following instrument adjustment and model deployment, the independent final sample ( $n = 30$ ) returned a mean compliance of 84.20% (SD = 15.30; median = 90.0%; range = 47–100%). The absolute difference was +37.87 percentage points (95% CI: 22.69–53.04).

The marked heteroscedasticity between phases — the pilot SD was 2.49 times the final SD — confirmed that a pooled-variance t-test was inappropriate. Welch’s t-test yielded  $t(38.11) = -5.05, p = 1.13 \times 10^{-5}$ . The effect size was Cohen’s *d* = 1.30 (Hedges’ *g* = 1.29; Glass’s  $\Delta$  = 0.99), consistent with a large-to-very-large effect by conventional benchmarks [13]. As illustrated in Figure 3, the distributional shift between phases was not merely a mean shift: the pilot distribution was strikingly bimodal and dispersed, with multiple licensees at 0% compliance, while the final distribution was substantially compressed and skewed toward high compliance. Table 3 presents the complete primary and secondary inferential statistics.

**Figure 2.** Comparative distributions of overall regulatory compliance scores in the pilot phase (Phase 3, n = 30) and final implementation phase (Phase 5, n = 30).

Box: IQR; horizontal line: median; whiskers:  $1.5 \times \text{IQR}$ ; diamonds: group means; points: individual observations.  
 Welch's t-test:  $t(38.11) = -5.05$ ,  $p = 1.13 \times 10^{-5}$ ; Cohen's  $d = 1.30$ ; Hedges'  $g = 1.29$ .

Recommended Tool: Python (Matplotlib/Seaborn) • Data/Variables to include: - Pilot phase (n=30): mean=46.33, SD=38.11, median=55.5, Q1=1.0, Q3=80.0, min=0, max=99 - Final phase (n=30): mean=84.20, SD=15.30, median=90.0, Q1=72.25, Q3=97.75, min=47, max=100 - Annotate: Welch's  $t(38.11) = -5.05$ ,  $p = 1.13 \times 10^{-5}$ , Cohen's  $d = 1.30$ , +37.87 pp (95% CI: 22.69–53.04) • Visual instructions: Single panel, two boxes side by side; y-axis 0–105%; colorblind-safe palette; significance bracket at top. [END BLUEPRINT]

**Table 3.** Primary and secondary inferential statistics comparing pilot and final implementation phases of the IVC model.

Indicator	Pilot phase (mean $\pm$ SD)	Final phase (mean $\pm$ SD)	Difference	95% CI	Welch's t	p	Cohen's d
Overall regulatory compliance (primary outcome)	46.33 $\pm$ 38.11	84.20 $\pm$ 15.30	+37.87 pp	22.69–53.04	-5.05	$1.13 \times 10^{-5}$	1.30
Major findings per visit	4.8 $\pm$ 1.7	1.4 $\pm$ 0.9	-3.4	—	—	—	—
Critical findings per visit	1.6 $\pm$ 0.8	0.5 $\pm$ 0.4	-1.1	—	—	—	—

• Recommended Tool: Tables (markdown/Word) • Rows: (1) Overall regulatory compliance — Pilot 46.33 $\pm$ 38.11, Final 84.20 $\pm$ 15.30, +37.87 pp, 95%CI 22.69–53.04,  $t(38.11) = -5.05$ ,  $p = 1.13 \times 10^{-5}$ ,  $d = 1.30$ ,  $g = 1.29$ ; (2) Major findings per visit — 4.8 $\pm$ 1.7 vs 1.4 $\pm$ 0.9, -3.4; (3) Critical findings per visit — 1.6 $\pm$ 0.8 vs 0.5 $\pm$ 0.4, -1.1. • Visual instructions: APA/Q1 horizontal-lines-only format; bold the overall compliance row. [END BLUEPRINT]

Note. Results are based on 60 field inspections: pilot phase (Phase 3,  $n = 30$ ) and final implementation phase (Phase 5,  $n = 30$ ). Welch's t-test with Satterthwaite-adjusted degrees of freedom was used given marked heteroscedasticity ( $SD_1 = 38.11$  vs.  $SD_2 = 15.30$ ; ratio = 2.49). Hedges'  $g = 1.29$ ; Glass's  $\Delta = 0.99$ . pp = percentage points; CI = confidence interval; — = not applicable.

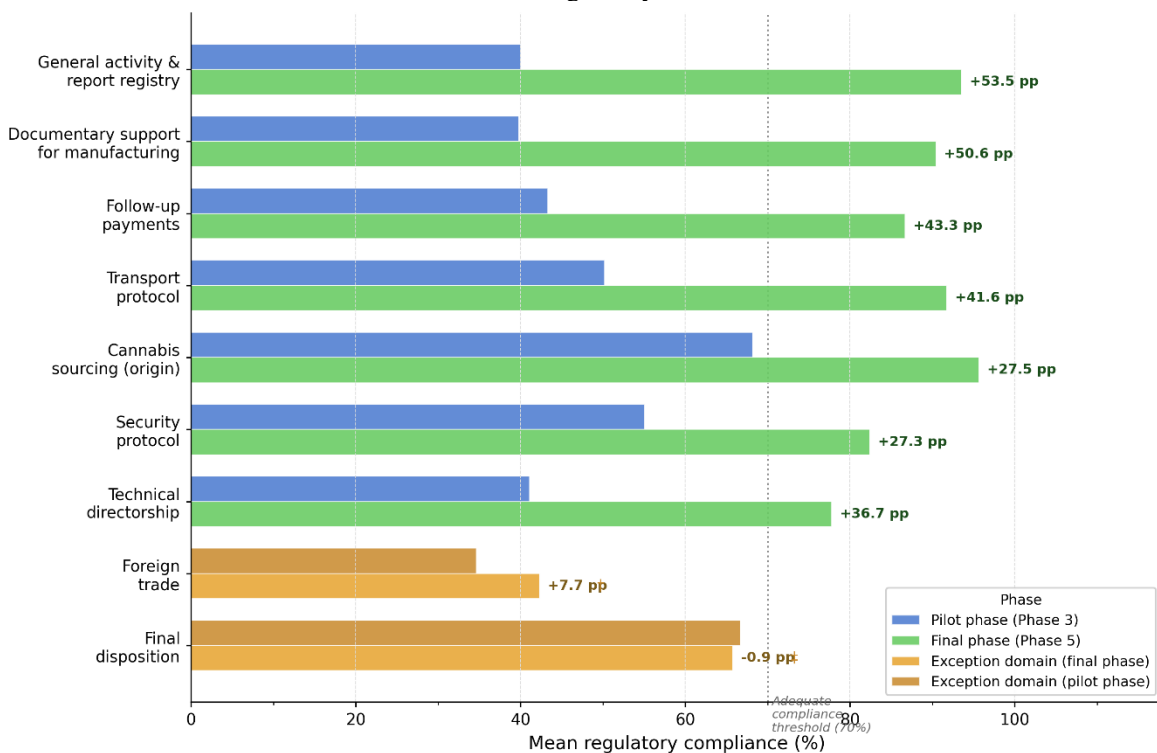
### 3.3 Domain-Level Compliance: Heterogeneous Gains Across Regulatory Axes

The aggregate improvement conceals a pattern of substantial heterogeneity across the nine scoreable regulatory axes. Four axes showed the most pronounced gains. Technical directorship compliance rose from a mean of 41.11% to 77.78%; follow-up payment compliance from 43.33% to 86.67%; documentary support for manufacturing from 39.81% to 90.42%; and the general activity and reporting registry from 40.04% to 93.54%. The median-level picture is sharper still: both technical directorship and follow-up payments had pilot-phase medians of 0% — meaning that in the typical licensee, these requirements produced no compliant responses whatsoever — rising to 100% in the final phase. Documentary support for manufacturing moved from a median of 0% to 88%, and the general registry from 26% to 95%.

Two domains showed stability rather than gain. Cannabis sourcing maintained a median of 100% across both phases, indicating that this axis was already well-managed before the model's introduction and functioned as an internal positive control for instrument validity. Foreign trade compliance, by contrast, improved only marginally — from a mean of 34.62% to 42.31% and a median that remained fixed at 50% in both phases — representing the only domain where the model's introduction did not produce a meaningful shift.

One domain declined: final disposition of controlled materials fell fractionally from 66.67% to 65.79%, a difference of less than one percentage point. Figure 4 displays the full domain-level compliance profile across both phases.

**Figure 4.** Axis-level mean regulatory compliance (%) in the pilot and final implementation phases across nine evaluated regulatory domains.



† Foreign trade: persistent low-compliance domain; minimal gain (+7.69 pp); median fixed at 50% in both phases.  
‡ Final disposition: only domain showing a decline (-0.88 pp). Both domains highlighted in amber.

• Recommended Tool: Python (Matplotlib/Seaborn) • Data/Variables to include (Pilot % → Final %, ordered by gain descending): - General activity and report registry: 40.04 → 93.54 (+53.50 pp) - Documentary support for manufacturing: 39.81 → 90.42 (+50.61 pp) - Transport protocol: 50.20 → 91.76 (+41.57 pp) - Follow-up payments: 43.33 → 86.67 (+43.34 pp) - Technical directorship: 41.11 → 77.78 (+36.67 pp) - Security protocol: 55.06 → 82.41 (+27.35 pp) - Cannabis sourcing (origin): 68.18 → 95.65 (+27.47 pp) - Foreign trade: 34.62 → 42.31 (+7.69 pp) — flag as persistent gap - Final disposition: 66.67 → 65.79 (-0.88 pp) — flag as only declining domain • Visual instructions: Paired horizontal grouped bar chart; reference line at 70%; Foreign trade and Final disposition bars in warning colour. [END BLUEPRINT]

### 3.4 Risk Classification Shift

Applying the study's pre-specified risk stratification thresholds — compliance below 50% = high risk; 50–70% = medium risk; above 70% = low risk — the distribution of licensees across risk tiers changed substantially between phases. In the pilot phase, 21 of 30 licensees (70.0%) were classified as high risk, 7 (23.3%) as medium

risk, and only 2 (6.7%) as low risk. In the final phase, the high-risk proportion fell to 13 of 30 (43.3%), medium risk remained at 7 (23.3%), and low-risk licensees increased to 10 (33.3%). Under the binary operationalisation of adequate compliance (> 70%), this translates to a relative risk of  $RR = 2.09$  (95% CI: 1.26–3.48) for achieving adequate compliance under the final model compared to the pilot baseline.

### 3.5 Sensitivity Analysis

The primary compliance effect withstood all pre-specified worst-case confounding scenarios. Replacing the highest 10% of final-phase scores with the pilot-phase mean reduced the estimated difference to +32.50 pp ( $p < 0.001$ ); replacing 20% yielded +27.27 pp ( $p = 0.001$ ); replacing 30% yielded +22.13 pp ( $p = 0.007$ ). The E-value for the point estimate of the binary outcome ( $RR = 2.09$ ) was 3.60, indicating that an unmeasured confounder would need to be associated with both phase allocation and compliance outcome by a factor of at least 3.6 to entirely explain the observed association [14]. The E-value for the lower confidence interval bound was 1.82.

### 3.6 Qualitative Findings and Triangulation

Thematic coding of the 12 semi-structured interviews and 2 focus groups produced four emergent categories, each with consistent frequency of mention across participants: greater clarity of evaluation criteria (83% of participants), better operational preparation for inspection visits (79%), reduction of inspector discretionality (75%), and an increased compliance culture among licensees (71%). These categories converged directly with the quantitative axes showing the largest gains. The axes where pilot-phase compliance was zero — technical directorship and follow-up payments — corresponded precisely to the qualitative theme of criterion ambiguity: participants described these requirements as previously unevaluated in practice, not because licensees were unwilling to comply, but because no standardised instrument had ever made the evaluation criteria explicit and actionable. The triangulation thus reinforces the primary quantitative finding while simultaneously offering a mechanism: the model's effect operated substantially through the clarity it introduced to an enforcement system that had previously functioned without a structured measuring instrument.

## 4. Discussion

### 4.1 Low Baseline Compliance Was Structural, Not Behavioural

The finding that commands the most interpretive attention is not the magnitude of the improvement — +37.87 percentage points, Cohen's  $d = 1.30$  — but the starting point that made such an improvement possible. A sector-wide mean compliance of 46.33%, with a median of just 55.5% and multiple licensees at absolute zero across several regulatory axes, is a profile that regulators in mature cannabis markets would find difficult to imagine. The question is why.

The qualitative evidence answers it directly. Eighty-three percent of regulatory actors and licensees cited the absence of clear evaluation criteria as the primary systemic weakness before the IVC model existed. This is consistent with a well-established pattern in first-generation regulatory frameworks: the legislative architecture arrives years ahead of the enforcement infrastructure. Canada experienced precisely this trajectory [1]; its 2014 shift to licensed commercial producers created compliance obligations before the monitoring machinery capable of verifying them was in place, generating what its own regulatory bodies described as a void that individual provincial colleges had to improvise to fill. Australia's experience was structurally parallel — the Therapeutic Goods Administration's expansion of prescribing access from 2016 onwards preceded the mandatory GMP quality standards that only came into force in 2023, leaving years of unverified production in between [2]. Colombia's situation differed only in the identity of the missing instrument: it was not quality standards that were absent, but the standardised inspection record through which those standards could be systematically evaluated. The 46.33% baseline, read this way, is not evidence of a non-compliant sector. It is evidence that a sector cannot demonstrate compliance with an instrument that does not yet exist.

This interpretation has an immediate policy implication. Regulators in emerging cannabis markets who interpret low compliance scores as a sign of actor resistance will draw the wrong conclusions and pursue the wrong interventions. The more accurate reading — supported by both the quantitative shift and the qualitative triangulation in this study — is that compliance behaviour is substantially shaped by the clarity and operationalisation of the measurement system itself.

### 4.2 Zero-Compliance Axes and the Mechanism of Instrument-Driven Change

The domain-level findings sharpen this argument. Technical directorship and follow-up payment compliance both registered median scores of 0% in the pilot phase — not low compliance, but the complete absence of measurable engagement with these regulatory dimensions. Both moved to 100% in the final phase. Documentary support for manufacturing went from a median of 0% to 88%; the general activity and reporting registry from 26% to 95%. These are not incremental adjustments in an already functioning system; they are discontinuous shifts that point to a threshold effect: once a structured instrument makes a requirement explicit, visible, and evaluable by a named inspector, the incentive to comply changes categorically.

This mechanism — instruments as governance actors, not merely measuring tools — finds support in the quality indicator literature. Boulkedid et al. [8] demonstrated that Delphi-derived quality indicators, precisely because they encode expert consensus into a public and reproducible format, reshape the behaviour of the professionals and organisations they measure. The 92% Delphi consensus achieved in this study's validation procedure is not incidental; it explains why the instrument was perceived as authoritative. Seventy-five percent of participants

noted a reduction in inspector discretionality, and 79% reported better preparation for visits — both of which are downstream effects of an instrument whose criteria can no longer be contested on grounds of ambiguity.

A limitation must be acknowledged directly alongside this finding: the pilot and final samples covered different licensees across partially different departments. Although this design was deliberate — to avoid conflating licensee learning with instrument effects — it introduces the possibility that geography partially confounded phase. Cundinamarca and Antioquia, which concentrated 59% of all visits, showed persistently different compliance profiles in the pilot phase (89.50% and 29.70%, respectively), suggesting that geographic variation in sector maturity is real. The sensitivity analysis reduces but does not eliminate this concern. Future research should track the same licensees longitudinally over repeated inspection cycles to isolate instrument effects from territorial dynamics.

#### **4.3 The Foreign Trade Exception Reveals Limits That Instruments Alone Cannot Resolve**

One domain resisted the general pattern entirely. Foreign trade compliance shifted from 34.62% to 42.31% — a marginal gain that left the median fixed at 50% across both phases. This is not a failure of the instrument to measure correctly; it is evidence of something the instrument correctly identified as a structural deficit that lies beyond the reach of any inspection model. As the source documents reveal, many licensees either do not conduct export activities at all — a reflection of the nascent state of Colombia's cannabis export market — or conduct them through informal postal channels that bypass official customs and FNE reporting obligations entirely, driven by the fragmented, magistral-preparation-oriented character of much of the sector's current international activity. The FNE's own trade data confirm the pattern: export authorisations grew from 156 in 2023 to 360 in 2024, but the proportion of licensees with compliant export documentation remained persistently low. This points to a coordination gap between the FNE and the national customs authority (DIAN) that no inspection record can bridge alone. The Australia-parallel here is instructive: Lim & Lim [2] documented precisely this failure mode — inter-agency coordination gaps that allowed high-volume prescribing to proceed without shared data across the TGA and the AHPRA — and concluded that regulatory harmonisation across agencies must be addressed legislatively, not operationally. Colombia's foreign trade compliance deficit warrants the same conclusion: a formal data-sharing protocol between the FNE and DIAN, mandated at regulatory level, is the intervention this evidence calls for.

#### **4.4 Instrument Standardisation as a Public Health Governance Tool**

The broader implication of this study concerns what inspection standardisation accomplishes beyond compliance measurement. The global pharmacovigilance literature has consistently shown that post-market safety monitoring of cannabis-based medicines is fragmented, poorly integrated into clinical workflows, and resistant to the kind of systematic surveillance that generates actionable regulatory intelligence [6]. Twenty-four monitoring systems identified across multiple jurisdictions share no common terminology, no unified reporting architecture, and produce data that are largely inaccessible in real time. The IVC model developed here operates at the manufacturing end of the same chain — not at the patient-safety end — but the structural problem is identical: the absence of a standardised instrument produces invisible non-compliance, which in turn generates diversion risk and erodes confidence in the regulated status of the product.

Framed this way, Colombia's contribution to the international literature is not merely a methodological one. It is a proof of concept: a middle-income country with a young regulatory apparatus was able to design, validate through expert consensus, deploy at scale across 12 departments, and evaluate a risk-based inspection instrument within a single implementation cycle, achieving large-magnitude compliance gains that withstood quantitative sensitivity testing. The E-value of 3.60 means that an unmeasured confounder would need to be nearly four times more prevalent among final-phase licensees and nearly four times more predictive of compliance to explain the observed effect away — a threshold that no plausible territorial or sector-maturity variable approaches [14]. The sample size ( $n = 30$  per phase), adequate for detecting large effects, does not support subgroup analyses by licence type, geographic region, or psychoactive versus non-psychoactive classification. Those analyses, which would allow risk stratification at a finer grain, constitute the immediate priority for the next phase of this research programme.

### **5. Conclusion**

This study provides the first empirically validated evidence that a structured, risk-based IVC model can produce statistically significant and operationally sustainable improvements in regulatory compliance among cannabis derivative manufacturers in a middle-income country context. Deployed across 60 field inspections in 12 Colombian departments, the model generated a mean compliance gain of +37.87 percentage points ( $d = 1.30$ ), driven by instrument-mediated changes in how regulatory requirements were communicated, evaluated, and internalised by licensees. The qualitative findings confirm that the instrument functions as a governance mechanism — reshaping actor behaviour through criterion clarity and inspector accountability — rather than merely as a passive measurement tool. The one persistent exception, foreign trade compliance, maps onto an inter-agency coordination gap that requires legislative rather than operational intervention. For regulators in Colombia and analogous emerging cannabis markets across Latin America and beyond, the study offers three concrete and time-bounded recommendations: first, institutionalise the IVC instrument as the mandatory inspection standard for all FNE oversight cycles within 12 months; second, establish a formal data-sharing protocol between the FNE and DIAN to close the foreign trade compliance gap within 24 months; and third, commission a longitudinal

inspection registry that tracks the same licensees over successive cycles, enabling risk-stratified oversight and generating the comparative evidence base that the sector's continued growth demands.

## References

- [1] M. Shim, H. Nguyen, P. Grootendorst, Lessons from 20 years of medical cannabis use in Canada, *PLoS ONE* 18 (2023) e0271079. <https://doi.org/10.1371/journal.pone.0271079>
- [2] E.C.N. Lim, C.E.D. Lim, Evolving health policy and regulatory oversight of medicinal cannabis in Australia: lessons for sustainable integration, *J. Cannabis Res.* 8 (2026) 31. <https://doi.org/10.1186/s42238-026-00394-z>
- [3] G.A. Calderón Vallejo, L.M. Pareja Hincapié, C. Caicedo Cano, R.A. Chica Ríos, Regulación del uso de marihuana en Colombia con fines medicinales, *Hacia Promoc. Salud* 22 (2017) 43–55. <https://doi.org/10.17151/hpsal.2017.22.1.4>
- [4] P.A. Cubillos Sánchez, Cannabis for medical and scientific purposes: the Colombian landscape, *Colomb. J. Anesthesiol.* 49 (2021) e954. <https://doi.org/10.5554/22562087.e954>
- [5] J.M. Orjuela-Rojas, X. García Orjuela, S. Ocampo Serna, Medicinal cannabis: knowledge, beliefs, and attitudes of Colombian psychiatrists, *J. Cannabis Res.* 3 (2021) 26. <https://doi.org/10.1186/s42238-021-00083-z>
- [6] R.Q. Wang, Y.A. Bonomo, C.M. Hallinan, Overview of global monitoring systems for the side effects and adverse events associated with medicinal cannabis use: a scoping review using a systematic approach, *BMJ Open* 14 (2024) e085166. <https://doi.org/10.1136/bmjopen-2024-085166>
- [7] Y. Zolotov, L. Mendoza Temple, R. Isralowitz, D.A. Gorelick, R. Abraham, D.I. Abrams, et al., Developing Medical Cannabis Competencies: a consensus statement, *JAMA Netw. Open* 8 (2025) e2535049. <https://doi.org/10.1001/jamanetworkopen.2025.35049>
- [8] R. Boulkedid, H. Abdoul, M. Loustau, O. Sibony, C. Alverti, Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review, *PLoS ONE* 6 (2011) e20476. <https://doi.org/10.1371/journal.pone.0020476>
- [9] P.N. Albers, C. Rinaldi, H. Brown, K.E. Mason, K. d'Apice, E. McGill, et al., Natural experiments for the evaluation of place-based public health interventions: a methodology scoping review, *Front. Public Health* 11 (2023) 1192055. <https://doi.org/10.3389/fpubh.2023.1192055>
- [10] J.Y. Chow, L. Geng, S. Bansal, B.S.L. Dickens, L.C. Ng, A.A. Hofmann, J.T. Lim, Evaluating quasi-experimental approaches for estimating epidemiological efficacy of non-randomised field trials: applications in Wolbachia interventions for dengue, *BMC Med. Res. Methodol.* 24 (2024) 170. <https://doi.org/10.1186/s12874-024-02291-6>
- [11] L. Kennedy-Shaffer, Quasi-experimental methods for pharmacoepidemiology: difference-in-differences and synthetic control methods with case studies for vaccine evaluation, *Am. J. Epidemiol.* 193 (2024) 1050–1058. <https://doi.org/10.1093/aje/kwae019>
- [12] I. Huh, J. Gim, Exploration of Likert scale in terms of continuous variable with parametric statistical methods, *BMC Med. Res. Methodol.* 25 (2025) 218. <https://doi.org/10.1186/s12874-025-02668-1>
- [13] C.R. Brydges, Effect size guidelines, sample size calculations, and statistical power in gerontology, *Innov. Aging* 3 (2019) igz036. <https://doi.org/10.1093/geroni/igz036>
- [14] I. Díaz, H. Lee, E. Kiciman, E.J. Schenck, M. Akacha, D. Follman, D. Ghosh, Sensitivity analysis for causality in observational studies for regulatory science, *J. Clin. Transl. Sci.* 7 (2023) e267. <https://doi.org/10.1017/cts.2023.688>
- [15] J. Xu, Y. Wang, Q. He, S. Xie, S. Feng, X. Wang, W. Wang, X. Sun, Evaluating the agreement between sensitivity and primary analyses in observational studies using routinely collected healthcare data: a meta-epidemiology study, *BMC Med.* 23 (2025) 393. <https://doi.org/10.1186/s12916-025-04199-4>