HIDDEN VALLEY
INTEGRITY OF THE GAME

Comprehensive Doctoral Executive Summary
of Research Methodology

2001-2026 | Oklahoma USSSA Baseball & Softball

20,316 Division-Level Records | 62 Data Fields | AppSheet Analytical Environment

Research Framework
Longitudinal Reconstruction | Economic Modeling | Public Accountability | Data Validation

Methodology Transparency Edition

Executive Abstract

Hidden Valley Integrity of the Game (HVIOG) is a longitudinal investigative and quantitative research initiative examining the structure, scale, economics, facility utilization, organizational relationships, and public-accountability implications of organized youth baseball and softball tournament activity in Oklahoma. The study spans 2001-2026 and evolved from event documentation and spreadsheet recordkeeping into a structured research database and custom Google AppSheet analytical environment. The current dataset contains 20,316 division-level records across 62 fields.

The project is not designed merely to produce a large statewide economic number. Its purpose is to reconstruct a historically fragmented youth-sports marketplace and transform dispersed records into a longitudinal dataset. The governing chain is documented source facts → researcher-coded variables → calculations → modeled estimates → aggregated findings → analytical interpretation.

1. Research Purpose and Design

The principal question is: What measurable economic activity, participation, facility utilization, organizational activity, and public-resource interaction can be reconstructed from organized USSSA baseball and softball tournament activity in Oklahoma between 2001 and 2026?

HVIOG is best characterized as a longitudinal, retrospective, observational, archival, data-reconstruction and economic-modeling study incorporating investigative and public-accountability methods. The objective is not simply to count tournaments; it is to reconstruct an economic and institutional ecosystem.

2. Unit of Analysis

The current dataset contains 20,316 division-level records, not necessarily 20,316 separate tournaments. The principal unit is generally a tournament-event/division record. A tournament with five divisions may generate five analytical records because divisions can have different team counts, fees, classifications, game requirements, umpire requirements, and economic calculations. Event counts, division records, team participation, and economic activity must therefore be reported separately.

3. Data Acquisition, Reconstruction, and Provenance

Historical information was progressively identified, collected, standardized, and incorporated into a structured environment. Sources include official tournament listings, USSSA event information, dates, locations, facilities, divisions, team counts, published fees, municipal documents, leases, public records, organizational publications, archived web material, correspondence, and historical research files.

Significant variables should be classified as Source-Observed, Researcher-Coded, Calculated, Modeled/Estimated, or Analytical/Inferential. This prevents modeled estimates from being represented as audited financial facts.

4. Development of the 62-Field Database

The present dataset contains 20,316 records x 62 fields, representing more than 1.25 million potential field-level data points. Its workflow is collection → standardization → storage → calculation → validation → filtering → aggregation → analysis → reporting. A public technical data dictionary should identify each important field's definition, source, classification, data type, formula, missing-value rule, and limitations.

5. Development of the AppSheet Research Application

A custom Google AppSheet application was developed as the operational interface between the researcher and the structured dataset. AppSheet became part of the research infrastructure rather than merely a presentation layer.

The architecture is Source Evidence → Data Acquisition → Structured Division Record → Standardized Fields → AppSheet Interface → Automated Expressions → Search/Filtering → Quality Control → Aggregation → Charts/Reports → Public Interpretation.

AppSheet addressed standardization, automated calculation, record retrieval, scalability, and quality control. It allows movement between individual records and statewide aggregates while applying consistent computational logic.

6. TournamentID Architecture and Data Integrity

A researcher-generated TournamentID system organizes analytical records. Every division-level record requiring independent identification should possess an appropriately unique identifier. A 2026 quality-control review detected identifiers reused across separate events. The affected records were isolated, compared against tournament, division, date, facility, location, and official event identifiers, corrected, and revalidated.

The resulting methodological rule is: identifiers are auditable data; database integrity is tested rather than presumed. The correction protocol is detect → investigate → correct → document → revalidate.

7. Quantitative Validation and Number Crunching

Validation includes duplicate and blank identifier detection, duplicate event/division review, event-ID comparison, missing-value detection, negative-value review, date consistency, facility normalization, outlier analysis, formula verification, record-count reconciliation, and aggregate-total reconciliation.

HVIOG uses a bottom-up calculation methodology: division record → teams/fees/games/attendance variables → record calculation → event aggregation → facility aggregation → municipal aggregation → annual aggregation → statewide longitudinal analysis. This preserves traceability from statewide findings back to source-level records.

8. Principal Economic Models

Team Entry Fee Activity: Entry Fee Per Team x Teams Entered = Total Team Entry Fee Activity. Current calculated/model-associated activity is approximately $36.34 million.

Gate Activity: Daily Fans x Daily Admission Fee = Total Gate Admission Fee; multiplied by Total Event Days to produce Total Event Gate. Current modeled activity is approximately $68.23 million.

Concession Activity: approximately $34.11 million in concession-related modeled activity, subject to disclosed attendance and spending assumptions.

Umpire Activity: approximately $31.09 million in modeled/calculated umpire-fee activity; a principal relationship is Total Event Umpires x Umpire Pay Per Game.

Economic Impact: approximately $92.86 million is stored in TotalEconomicImpactPerEvent. Because economic impact may include direct, indirect, induced, visitor, or multiplier-based activity, every component must be explicitly defined before any combined statewide conclusion is published.

9. Interpretation, Missing Data, and Assumptions

A governing rule is: economic activity is not synonymous with profit. Modeled revenue is not necessarily actual cash receipt, and gross activity is not taxable income. Aggregate modeling alone cannot establish tax liability, fraud, financial misconduct, or individual enrichment.

HVIOG distinguishes unknown from zero. A formal Assumption Register should identify each material model assumption, its numerical value, evidentiary basis, period of applicability, sensitivity, and limitations. Materially uncertain variables should be tested under conservative, baseline, and upper-bound scenarios.

10. Evidence Hierarchy and Finding Classification

The recommended hierarchy is: primary governmental/official records; primary organizational records; contemporaneous supporting documentation; secondary reporting; witness/source information; and researcher calculations/models. Significant findings should be labeled Documented, Calculated, Modeled/Estimated, Inferred, or Unverified.

An anomaly is a research signal, not proof of misconduct. It should trigger source reexamination, alternative explanations, additional documentation, and appropriate evidentiary classification.

11. Reproducibility, Version Control, and Auditability

Every public dataset release should carry a version identifier, release date, row count, column count, research period, validation status, known limitations, and material corrections. A permanent correction ledger should preserve changes rather than silently overwrite research history.

The forward reproducibility chain is Source → Source Variable → Database Record → Field → Formula → Result → Aggregation → Published Finding. The reverse path should also be possible: Finding → Calculation → Record → Source. This bidirectional traceability is the project's auditability standard.

12. Bias Controls, Ethics, and Evidentiary Boundaries

HVIOG arises from concerns about integrity, transparency, economics, governance, public facilities, and accountability. Controls against confirmation bias include preserving contrary evidence, favoring primary sources, standardizing formulas, treating comparable records consistently, disclosing assumptions, separating observations from interpretations, correcting errors, documenting uncertainty, and revising findings when better evidence emerges.

The methodology must permit the evidence to contradict the researcher's original expectation.

Because the activity involves children, accountability should focus on systems, adult decision-makers, organizations, financial structures, governmental entities, policies, facilities, and governance. Personally identifying information concerning minors should be minimized where unnecessary.

13. Data, Analysis, Interpretation, and Allegation

HVIOG distinguishes four levels: Data - what the source establishes; Analysis - what calculations and comparisons demonstrate; Interpretation - reasonable conclusions derived from evidence; and Questions or Allegations Requiring External Determination - matters requiring review by competent authorities.

HVIOG is not, by itself, a governmental audit, tax determination, criminal investigation, judicial finding, or organizational disciplinary determination. The database should illuminate questions, not substitute for the authority legally responsible for answering them.

14. Public-Facility and Municipal Research

Facility-level analysis can evaluate tournament activity, teams, entry fees, gate activity, concessions, umpire activity, event-host activity, municipal ownership, public investment, leases, operating agreements, facility improvements, utilities, subsidies, concession rights, and other governmental relationships. The nested architecture is State → Municipality → Facility → Tournament/Event → Division → Analytical Record. This bridges sports economics with public-administration research.

15. Validity, Reliability, Triangulation, and Human Oversight

Internal validity requires accurate data and conceptually appropriate models. Reliability asks whether the same inputs and formulas produce the same outputs. External validity is bounded by the population studied: Oklahoma USSSA baseball and softball. Strong findings should use triangulation across database records, municipal documents, facility records, organizational sources, and other independent evidence.

AppSheet is an analytical instrument, not independent evidence. Source evidence establishes the input; AppSheet organizes and processes it; formulas transform it; the researcher interprets the output. Automation requires both computational and conceptual validation.

16. Methodological Evolution and Governance

HVIOG developed through four broad phases: Observation and Documentation; Data Reconstruction; Analytical Infrastructure; and Formalization, Validation, and Public Reproducibility. The recommended governance cycle is Acquire → Preserve → Standardize → Classify → Calculate → Validate → Review → Version → Analyze → Publish → Correct → Revalidate.

17. Recommended Public Transparency Package

The methodology should be published as a coordinated transparency package: (1) Doctoral Executive Summary of Research Methodology; (2) Technical Calculation Manual; (3) 62-Field Data Dictionary; (4) Assumption Register; and (5) Dataset Validation and Correction Log.

The future methodology landing page should visually communicate: Source Data → 20,316 Structured Records → 62-Field Database → AppSheet Research Application → Calculation Engine → Data-Integrity Audit → Facility/Municipal/Statewide Analysis → Public Research.

18. Conclusion

Hidden Valley Integrity of the Game represents an effort to reconstruct approximately twenty-five years of Oklahoma youth baseball and softball tournament activity from fragmented historical information and transform those records into a structured analytical environment. Its current architecture - 20,316 division-level records across 62 fields covering 2001-2026 - permits individual event information to be examined within a broader longitudinal framework.

The development of AppSheet was a methodological turning point, enabling standardized data entry, automated calculation, retrieval, filtering, comparative analysis, quality control, and scalable research. The TournamentID review demonstrates that data integrity is an ongoing process rather than a declaration of perfection.

The public methodological standard should remain: Follow the evidence. Preserve the source. Show the calculation. Disclose the assumption. Test the data. Document the correction. Separate estimates from facts. Distinguish questions from conclusions.

Methodology Transparency Statement

Hidden Valley Integrity of the Game is an independent research initiative. Economic figures identified as calculated, modeled, or estimated are research outputs derived from available source information, database variables, formulas, and disclosed assumptions. Unless specifically supported by independently audited records, such figures should not be interpreted as verified organizational revenue, profit, taxable income, or actual cash receipts. Identification of an anomaly, inconsistency, or unexplained economic relationship does not independently establish unlawful conduct. Findings should be evaluated according to their underlying evidence, methodology, assumptions, and stated limitations.