

Supplementary Material for:
*From ‘Six Others’ to edge conservation: A
network reconciliation of gambling harm to others*

This document contains complete methodological details in Section 1, supported by mathematical derivations and model specifications given in Section 2. Section 5 contains some notes on computational implementation, including a link to the code repository.

Contents

1	Detailed Methodology	2
1.1	Data Harmonisation and Anchors	2
1.1.1	Pooling Prevalence Anchors	2
1.1.2	Anchoring Gambler–AO Overlap	3
1.1.3	Adjusting Fan-out for Commensurability	3
1.2	Edge Conservation and the Core Identity	3
1.3	Reconciling Fan-In via a Mixture Model	4
1.3.1	The Role of AO–Gambler Overlap	4
1.4	Extreme Scenarios: Scaled Self-Report and AO-Proxy Report	4
1.5	Reconciling Fan-Out via Competing Blending Models	5
2	Mathematical Derivations and Model Specifications	6
2.1	Derivation of the Edge-Conservation Identity	6
2.2	Closed-Form Solution for the Blending Weight	7
2.2.1	Interior Solutions and Generalisability	7
2.3	Derivation of Extreme Scenarios	8
2.4	Value Blending vs. Ratio Blending	9
2.4.1	Model 1: Value Blending (Additive Bias)	9
2.4.2	Model 2: Ratio Blending (Multiplicative Bias)	9
2.5	Mixture Model for Fan-In Heterogeneity	9
2.5.1	Likelihood Specification	10
2.5.2	Constraints and Estimation	10
3	Detailed Empirical Results	10
3.1	Prevalence Study Characteristics and Meta-Analysis	10
3.2	Data Preparation and Raw Distributions	11

3.3 Mixture Model Diagnostics	13
3.4 Solving for the Blending Weight	14
4 Notation	17
5 Computational Implementation	17

1 Detailed Methodology

The estimation of gambling-related harm to AOs is complicated by the use of multiple, disparate data sources that provide conflicting estimates. Our study's primary methodological aim is to develop a framework to reconcile and fuse these data types, leveraging their complementary strengths to produce a single, mathematically consistent model.

We identify five key data types, each measuring different aspects of the harm network, and each with distinct strengths and weaknesses:

1. **Prevalence Anchors:** Unconditional prevalence of gambler risk levels and AOs from large, representative surveys. Highest-confidence node counts but no edge information.
2. **Fan-out (AO-Proxy):** Number of people harmed per gambler, reported by AOs. May over-attribute harm; lacks direct knowledge of all affected parties.
3. **Fan-out (Self-Report):** Number of people harmed, reported by gamblers. Direct knowledge but prone to under-reporting from stigma or lack of awareness.
4. **Fan-in:** Number of gamblers harming each AO, reported by AOs. Captures clustering and multiple harm exposures invisible in prevalence data alone.
5. **Gambler Severity (AO-Report):** PGSI risk level of the gambler identified by AOs as the source of harm (typically either the "closest" or "most harmful"). Subject to selection bias when AOs are harmed by multiple gamblers. Used here for external validation only.

1.1 Data Harmonisation and Anchors

To ensure all data are comparable, three harmonisation steps are performed.

1.1.1 Pooling Prevalence Anchors

We first establish our 'ground truth' population parameters by performing an inverse-variance weighted meta-analysis of recent, high-quality jurisdictional surveys. This yields pooled estimates for:

- p_{PG} , p_{MR} , p_{LR} : The prevalence of PG, MR and LR gamblers, based on the PGSI.
- p_{AO} : The prevalence of adult AOs.

These pooled estimates serve as hard anchors or constraints in our model. Define the prevalence of any-risk gamblers as

$$p_{G_{any}} \equiv p_{LR} + p_{MR} + p_{PG}, \quad (1)$$

where G_{any} denotes the set of any-risk gamblers ($\text{PGSI LR} \cup \text{MR} \cup \text{PG}$).

1.1.2 Anchoring Gambler–AO Overlap

In addition to unconditional prevalences p_{AO} and $p_{G_{\text{any}}}$ (Eq. 1), we harmonise the conditional overlap reported variously across studies as either $P(\text{AO} \mid G_{\text{any}})$ or $P(G_{\text{any}} \mid \text{AO})$. We reconcile these via Bayes’ theorem:

$$P(\text{AO} \mid G_{\text{any}}) = \frac{P(G_{\text{any}} \mid \text{AO}) p_{\text{AO}}}{p_{G_{\text{any}}}}, \quad P(G_{\text{any}} \mid \text{AO}) = \frac{P(\text{AO} \mid G_{\text{any}}) p_{G_{\text{any}}}}{p_{\text{AO}}}. \quad (2)$$

We compute the overall AO–gambler overlap as

$$\pi_{\text{overlap}} \equiv P(G_{\text{any}} \mid \text{AO}) = \sum_{s \in \{\text{LR}, \text{MR}, \text{PG}\}} \frac{P(\text{AO} \mid s) p_s}{p_{\text{AO}}}. \quad (3)$$

We meta-analytically pool these quantities alongside the prevalence anchors to obtain a pooled point estimate and CI for π_{overlap} , which will later constrain the mixture and propagate uncertainty into $\mu_{\text{in}}(\pi)$ and $\alpha(\pi)$.

1.1.3 Adjusting Fan-out for Commensurability

Because most AO prevalence anchors are adult-only, we calibrate raw fan-out reports to adult-equivalent values using a multiplicative factor f_{adult} derived from a simple demographic ratio of the adult population (18+) to the school-age-and-older (5+) population. The justification and calculation of this factor, along with robustness checks and validation against household-composition data, are presented in Section 3.2.

All raw fan-out estimates (μ_{raw}) are adjusted to produce an adult-only equivalent:

$$\mu_{\text{adult}} = \mu_{\text{raw}} \times f_{\text{adult}}. \quad (4)$$

1.2 Edge Conservation and the Core Identity

The reconciliation framework rests on a fundamental conservation principle: every harm relationship can be counted either from the source (gamblers originating harm) or from the destination (AOs receiving harm). Let E_{total} denote the total number of directed harm edges in the population. By definition, these two counts must be identical:

$$E_{\text{total}} = \sum_{i \in G} \deg_{\text{out}}(i) = \sum_{j \in \text{AO}} \deg_{\text{in}}(j).$$

Partitioning the gambler population G into risk categories $r \in \{\text{LR}, \text{MR}, \text{PG}\}$ with unconditional prevalences p_r , and denoting mean fan-out as $\mu_{\text{out},r}$, the edges originating from gamblers are $N \sum_r p_r \mu_{\text{out},r}$. Similarly, with AO prevalence p_{AO} and mean fan-in μ_{in} , edges terminating at AOs are $N p_{\text{AO}} \mu_{\text{in}}$. Equating these yields the *core conservation identity*:

$$\sum_{r \in \{\text{LR}, \text{MR}, \text{PG}\}} p_r \mu_{\text{out},r} = p_{\text{AO}} \mu_{\text{in}}. \quad (5)$$

This identity is definitional: it follows from counting the same directed harm relationships in two ways. It assumes (i) a closed population and (ii) zero-risk gamblers ($\text{PGSI} = 0$) originate no harm edges. The full derivation appears in Section 2.1.

1.3 Reconciling Fan-In via a Mixture Model

The empirical fan-in distribution exhibits a sharp mode at $k = 1$ (67% of AOs report harm from exactly one gambler) but a heavy right tail incompatible with single-parameter count models. This implies the AO population is a mixture of two distinct subpopulations with different network contexts.

We model this as a two-component mixture constrained by the meta-analytic overlap parameter $\pi_{\text{overlap}} = P(G_{\text{any}} \mid \text{AO})$. Type 1 (*non-overlap*) AOs are not at-risk gamblers themselves; their fan-in follows a zero-truncated Poisson concentrated at $k \approx 1$. Type 2 (*overlap*) AOs are also at-risk gamblers, embedded in denser harm networks with heavier-tailed fan-in modelled by a zero-truncated negative binomial. The overall mean fan-in is therefore

$$\mu_{\text{in}}(\pi_{\text{overlap}}) = (1 - \pi_{\text{overlap}})\mu_{\text{non-overlap}} + \pi_{\text{overlap}}\mu_{\text{overlap}}. \quad (6)$$

Component parameters are estimated via maximum likelihood to reproduce the pooled empirical distribution (see Section 2.5 for full specification and diagnostic results). This formulation captures clustering without assuming uniform exposure, revealing that the overlap subgroup (meta-analytically fixed at 24.5% of AOs) experiences substantially higher fan-in ($\mu_{\text{overlap}} \approx 2.7$ vs. $\mu_{\text{non-overlap}} \approx 1.3$).

1.3.1 The Role of AO–Gambler Overlap

The overlap parameter π_{overlap} (Eq. 3) plays a central role in reconciliation because it quantifies the clustering of harms within social networks. By governing the weight of the heavy-tailed component in the mixture model (Eq. 6), π_{overlap} directly determines the population mean fan-in (μ_{in}). Because the conservation identity (Eq. 5) requires edges from gamblers to equal edges received by AOs, a higher overlap implies a denser fan-in network, increasing the total edge count required to satisfy conservation.

Consequently, π_{overlap} acts as a lever on the blending weight α : a larger overlap increases the required edge budget, shifting the solution for α toward the higher counts typical of problem gambler self-reports. Conversely, lower overlap implies sparser networks, shifting α toward AO-proxy reports. In this way, the empirically observed clustering anchors the reconciliation of conflicting fan-out data.

1.4 Extreme Scenarios: Scaled Self-Report and AO-Proxy Report

Before blending the conflicting fan-out reports, we first establish the bounds of the reconciliation problem by calculating two extreme scenarios. Each scenario assumes one raw report has the correct *relative* harm profile but requires uniform scaling to satisfy the conservation identity.

Let μ_{target} denote either μ_{self} or $\mu_{\text{ao-proxy}}$. We solve for a scalar scaling factor s such that the scaled vector $s \cdot \mu_{\text{target}}$ satisfies Equation (5):

$$s = \frac{p_{\text{AO}} \mu_{\text{in}}(\pi_{\text{overlap}})}{\sum_r p_r \mu_{\text{target},r}}. \quad (7)$$

We compute this twice: once using μ_{self} (the ‘Self-Only’ scenario) and once using $\mu_{\text{ao-proxy}}$ (the ‘AO-Only’ scenario). These provide conservation-consistent fan-out estimates that represent the extreme bounds of our subsequent blending models (full derivation in Section 2.3).

1.5 Reconciling Fan-Out via Competing Blending Models

The extreme scenarios assume only one report’s relative shape is correct. A more plausible approach assumes both reports contain idiosyncratic bias and seeks a blend most consistent with the anchor data. We solve for a single blending weight α that makes the convex combination of the two reports satisfy the conservation identity:

$$\sum_r p_r [(1 - \alpha)\mu_{\text{ao-proxy},r} + \alpha\mu_{\text{self},r}] = p_{\text{AO}} \mu_{\text{in}}(\pi_{\text{overlap}}). \quad (8)$$

This linear equation yields a unique closed-form solution (see Section 2.2). We note that a valid interior solution ($\alpha \in [0, 1]$) is guaranteed only because the target edge count derived from prevalence anchors falls strictly between the totals implied by the raw self-report and AO-proxy sources. This confirms that the conflicting reports effectively bracket the population-consistent truth.

The solved weight α parameterises two reconciliation models reflecting different bias theories:

- **Model 1 (Value Blend)** assumes additive bias—errors roughly constant across severity levels—yielding the component-wise convex combination in Equation (8).
- **Model 2 (Ratio Blend)** assumes multiplicative bias—errors proportional to magnitude—and blends log-ratios before uniform scaling (Section 2.4).

Both use the same α^* and satisfy conservation, but embody different measurement assumptions.

We report reconciled fan-out estimates without confidence intervals. Bootstrap CIs accounting for sampling error in the prevalence and fan-in inputs yield standard errors of 0.05–0.11, but these substantially understate true uncertainty. The primary source of uncertainty is structural—whether reporting bias is additive (Value Blend) or multiplicative (Ratio Blend)—which produces a range of 1.8 units for problem-gambler fan-out that cannot be reduced by larger samples. Presenting both models with their distinct assumptions more accurately conveys the state of knowledge than interval estimates that would imply false precision.

2 Mathematical Derivations and Model Specifications

2.1 Derivation of the Edge-Conservation Identity

We treat the population as a finite, closed network where directed edges represent gambling-related harm relationships. Let N denote the total adult population size. The set of any-risk gamblers G is partitioned into three disjoint risk categories $r \in \{\text{LR}, \text{MR}, \text{PG}\}$ with unconditional prevalences p_r . For each gambler $i \in G_r$, let $\deg_{\text{out}}(i)$ denote the number of AOs they harm (their fan-out). The group-level mean fan-out is

$$\mu_{\text{out},r} = \frac{1}{|G_r|} \sum_{i \in G_r} \deg_{\text{out}}(i), \quad (9)$$

where $|G_r| = Np_r$ is the number of gamblers in category r .

The total number of harm edges originating from gamblers is therefore

$$\begin{aligned} E_{\text{total}} &= \sum_{r \in \{\text{LR}, \text{MR}, \text{PG}\}} \sum_{i \in G_r} \deg_{\text{out}}(i) \\ &= \sum_r (Np_r) \mu_{\text{out},r} \\ &= N \sum_r p_r \mu_{\text{out},r}. \end{aligned} \quad (10)$$

Similarly, let AO denote the set of AOs with unconditional prevalence p_{AO} and size $|\text{AO}| = Np_{\text{AO}}$. For each AO j , let $\deg_{\text{in}}(j)$ denote the number of gamblers harming them (their fan-in). The mean fan-in is

$$\mu_{\text{in}} = \frac{1}{|\text{AO}|} \sum_{j \in \text{AO}} \deg_{\text{in}}(j). \quad (11)$$

Counting edges from the destination side gives

$$E_{\text{total}} = |\text{AO}| \mu_{\text{in}} = Np_{\text{AO}} \mu_{\text{in}}. \quad (12)$$

Since each harm relationship is a single directed edge, the totals in (10) and (12) must be identical. Equating them and dividing by N yields the *core conservation identity*:

$$\sum_{r \in \{\text{LR}, \text{MR}, \text{PG}\}} p_r \mu_{\text{out},r} = p_{\text{AO}} \mu_{\text{in}}. \quad (13)$$

This identity is definitional: it follows from counting the same set of edges in two ways. We assume (i) the population is closed with respect to gambling-related harm, and (ii) zero-risk gamblers ($\text{PGSI} = 0$) originate no harm edges.

2.2 Closed-Form Solution for the Blending Weight

We model the true fan-out vector $\boldsymbol{\mu}_{\text{out}}(\alpha)$ as a convex combination of the two conflicting reports:

$$\boldsymbol{\mu}_{\text{out}}(\alpha) = (1 - \alpha) \cdot \boldsymbol{\mu}_{\text{ao-proxy}} + \alpha \cdot \boldsymbol{\mu}_{\text{self}}, \quad (14)$$

where $\alpha \in [0, 1]$ is the weight on gambler self-reports, $\boldsymbol{\mu}_{\text{ao-proxy}} = (\mu_{\text{ao,LR}}, \mu_{\text{ao,MR}}, \mu_{\text{ao,PG}})$, and $\boldsymbol{\mu}_{\text{self}} = (\mu_{\text{self,LR}}, \mu_{\text{self,MR}}, \mu_{\text{self,PG}})$.

Substituting (14) into the conservation identity (13) yields a single equation in one unknown α :

$$\sum_r p_r [(1 - \alpha)\mu_{\text{ao},r} + \alpha\mu_{\text{self},r}] = p_{\text{AO}} \mu_{\text{in}}(\pi_{\text{overlap}}). \quad (15)$$

Define the total per-capita edges implied by each raw source:

$$E_{\text{ao-proxy}} \equiv \sum_r p_r \mu_{\text{ao},r}, \quad (16)$$

$$E_{\text{self}} \equiv \sum_r p_r \mu_{\text{self},r}, \quad (17)$$

$$E_{\text{target}} \equiv p_{\text{AO}} \mu_{\text{in}}(\pi_{\text{overlap}}). \quad (18)$$

Expanding (15) and using (16)–(18) gives a linear equation in α :

$$(1 - \alpha)E_{\text{ao-proxy}} + \alpha E_{\text{self}} = E_{\text{target}}. \quad (19)$$

Rearranging yields a unique closed-form solution, provided $E_{\text{self}} \neq E_{\text{ao-proxy}}$:

$$\alpha^* = \frac{E_{\text{target}} - E_{\text{ao-proxy}}}{E_{\text{self}} - E_{\text{ao-proxy}}}. \quad (20)$$

The solution exists in $[0, 1]$ if and only if E_{target} lies between the two extreme edge counts:

$$\min(E_{\text{ao-proxy}}, E_{\text{self}}) \leq E_{\text{target}} \leq \max(E_{\text{ao-proxy}}, E_{\text{self}}). \quad (21)$$

In our data, $E_{\text{self}} < E_{\text{target}} < E_{\text{ao-proxy}}$, so $\alpha^* \in (0, 1)$ is uniquely determined. The conditions under which this interior solution exists, and principled alternatives when it does not, are discussed further below in section 2.2.1.

2.2.1 Interior Solutions and Generalisability

The closed-form solution for α^* exists in the interior of $[0, 1]$ if and only if the target edge count E_{target} is bracketed by the raw fan-out reports:

$$\min(E_{\text{ao-proxy}}, E_{\text{self}}) < E_{\text{target}} < \max(E_{\text{ao-proxy}}, E_{\text{self}}). \quad (22)$$

In the present application, this condition holds: $E_{\text{self}} = 0.042$, $E_{\text{ao-proxy}} = 0.184$, and $E_{\text{target}} = 0.110$ (per-capita edges, using pooled prevalence anchors and adult-adjusted fan-out). Consequently, $\alpha^* = 0.52$ uniquely solves the conservation identity as a convex combination of the two reports. This interior solution is non-trivial: it emerges because the evidence hierarchy adopted here treats prevalence and fan-in estimates as higher-confidence anchors, while the two fan-out sources exhibit *opposing* biases that straddle the true population quantity.

The framework presented here is tailored to the specific data structure of the present application: two fan-out sources with opposing biases, high-quality prevalence anchors, and an interior solution satisfying (22). When these conditions are not met, the core reconciliation principle—that network identities must constrain measurement—admits natural extensions. If the bracketing condition fails (e.g., all sources share directional bias relative to E_{target}), one must either relax the evidence hierarchy by treating some anchors as uncertain parameters within a joint estimation framework, or accept a boundary solution ($\alpha^* \in \{0, 1\}$) and apply uniform scaling as in section 2.3. If more than two fan-out sources are available, the blending weight generalises to a vector $\mathbf{w}$ solved via precision-weighted pooling subject to the conservation constraint. The present hard-constraint approach is warranted because large-scale prevalence surveys provide robust node counts and the interior compromise reconciles genuine measurement conflict rather than merely splitting arbitrary differences.

2.3 Derivation of Extreme Scenarios

Before blending, we establish bounds by assuming one raw report has the correct *relative* harm profile but requires uniform scaling. Let μ_{target} denote either μ_{self} or $\mu_{\text{ao-proxy}}$. The true fan-out is proportional to this target:

$$\mu_{\text{out}} = s \cdot \mu_{\text{target}}, \quad (23)$$

where s is a scalar scaling factor.

Substituting (23) into the conservation identity (13) gives

$$s \sum_r p_r \mu_{\text{target},r} = p_{\text{AO}} \mu_{\text{in}}(\pi_{\text{overlap}}). \quad (24)$$

Using the definitions (16)–(18), the scaling factor is

$$s = \frac{E_{\text{target}}}{E_{\text{target, raw}}}, \quad (25)$$

where $E_{\text{target, raw}}$ is the unscaled edge count from the chosen source. We compute two scenarios:

- **Self-Only:** $\mu_{\text{target}} = \mu_{\text{self}}$, giving $s_{\text{self}} = E_{\text{target}}/E_{\text{self}}$
- **AO-Only:** $\mu_{\text{target}} = \mu_{\text{ao-proxy}}$, giving $s_{\text{ao}} = E_{\text{target}}/E_{\text{ao-proxy}}$

These scenarios preserve the raw severity gradient while enforcing conservation through uniform scaling.

2.4 Value Blending vs. Ratio Blending

The solved weight α^* is used to parameterise two competing theories of reporting bias.

2.4.1 Model 1: Value Blending (Additive Bias)

Under an additive bias model, the true fan-out is the component-wise convex combination:

$$\mu_{\text{blend},r} = (1 - \alpha^*) \mu_{\text{ao},r} + \alpha^* \mu_{\text{self},r}, \quad r \in \{\text{LR}, \text{MR}, \text{PG}\}. \quad (26)$$

This preserves the conservation identity by construction and minimises squared deviation in absolute fan-out values.

2.4.2 Model 2: Ratio Blending (Multiplicative Bias)

Under a multiplicative bias model, the *relative shape* of harm across severity levels is more reliable than absolute numbers. We blend log-ratios relative to a reference group (LR gamblers):

$$r_{\text{ao},r} = \log\left(\frac{\mu_{\text{ao},r}}{\mu_{\text{ao},\text{LR}}}\right), \quad r \in \{\text{MR}, \text{PG}\}, \quad (27)$$

$$r_{\text{self},r} = \log\left(\frac{\mu_{\text{self},r}}{\mu_{\text{self},\text{LR}}}\right), \quad r \in \{\text{MR}, \text{PG}\}. \quad (28)$$

The blended log-ratio vector is

$$r_{\text{blend},r} = (1 - \alpha^*) r_{\text{ao},r} + \alpha^* r_{\text{self},r}. \quad (29)$$

Exponentiating yields the shape vector, which is then uniformly scaled using the method of section 2.3 to satisfy conservation. The final fan-out values are

$$\mu_{\text{blend},r} = s_{\text{ratio}} \cdot \mu_{\text{base}} \cdot \exp(r_{\text{blend},r}), \quad (30)$$

where μ_{base} is the blended LR value and s_{ratio} is the scaling factor ensuring the conservation identity holds.

2.5 Mixture Model for Fan-In Heterogeneity

Let K_j denote the fan-in (number of gamblers harming) for AO j . The fan-in distribution is modelled as a two-component mixture with fixed weight π_{overlap} (from meta-analysis) and variable component parameters:

$$K_j \sim \begin{cases} \text{ZTP}(\mu_{\text{non-overlap}}), & \text{with probability } 1 - \pi_{\text{overlap}}, \\ \text{ZTNB}(\mu_{\text{overlap}}, \theta), & \text{with probability } \pi_{\text{overlap}}, \end{cases} \quad (31)$$

where ZTP is zero-truncated Poisson and ZTNB is zero-truncated negative binomial with dispersion θ . The negative binomial component allows for overdispersion in the overlap group, reflecting the greater clustering of harm in dense gambling networks.

2.5.1 Likelihood Specification

Let $\mathbf{k} = \{k_j\}_{j=1}^{N_{\text{AO}}}$ be the observed fan-in counts across the $N_{\text{AO}} = 1,819$ pooled AO sample, with empirical frequencies n_k for $k = 1, 2, 3, 4$ and n_{5+} for $k \geq 5$. The multinomial log-likelihood is

$$\ell(\phi; \mathbf{k}) = \sum_{k=1}^4 n_k \log P(k; \phi) + n_{5+} \log P(k \geq 5; \phi), \quad (32)$$

where $\phi = (\mu_{\text{non-overlap}}, \mu_{\text{overlap}}, \theta)$ and the mixture probability mass function is

$$P(k; \phi) = (1 - \pi_{\text{overlap}}) f_{\text{ZTP}}(k; \mu_{\text{non-overlap}}) + \pi_{\text{overlap}} f_{\text{ZTNB}}(k; \mu_{\text{overlap}}, \theta). \quad (33)$$

The right-tail probability is $P(k \geq 5; \phi) = 1 - \sum_{k=1}^4 P(k; \phi)$.

2.5.2 Constraints and Estimation

The optimisation is subject to three constraints derived from empirical evidence:

$$\pi_{\text{overlap}} = 0.2452 \quad (\text{meta-analytic estimate}), \quad (34)$$

$$\mu_{\text{overlap}} \geq \mu_{\text{non-overlap}} \quad (\text{network density ordering}), \quad (35)$$

$$\mu_{\text{non-overlap}} > 0 \quad (\text{positivity}). \quad (36)$$

The implied population mean fan-in is

$$\hat{\mu}_{\text{in}} = (1 - \pi_{\text{overlap}}) \hat{\mu}_{\text{non-overlap}} + \pi_{\text{overlap}} \hat{\mu}_{\text{overlap}}, \quad (37)$$

which is then substituted into the conservation identity (13) for reconciliation. All numerical optimisation was performed in Python using SciPy's SLSQP algorithm to maximise (32) subject to constraints (34)–(36); see the code archive for implementation details.

3 Detailed Empirical Results

3.1 Prevalence Study Characteristics and Meta-Analysis

First, key parameters from six recent, large-scale Australian jurisdictional surveys were collated. This list was compiled after identifying all recent (since 2017) Australian prevalence studies and then selecting those that reported the three key parameters required for this reconciliation model: the prevalence of gambler risk levels, the prevalence of AOs, and the conditional overlap between these two groups. These parameters serve as the primary inputs for the subsequent stages of the model.

Table 1 summarises the source-study characteristics most relevant to interpreting these estimates, including the AO screening or inclusion wording where it was available in the source document. Across studies using PGSI categories, standard scoring bands were used: non-problem gambling = 0, low-risk gambling = 1–2, moderate-risk gambling = 3–7, and problem gambling = 8 or above. Table 2 then summarises the quantitative inputs, detailing sample sizes (N) alongside the prevalence of AOs (p_{AO}) and specific gambler risk levels ($p_{\text{LR}}, p_{\text{MR}}, p_{\text{PG}}$). The table also presents the derived any-risk prevalence (p_{Gany}) and the overlap parameter (π_{overlap}). We pooled estimates from the six surveys (Table 2) using an inverse-variance weighted meta-analysis. Given the structural importance of the gambler–AO overlap, Table 3 details the derivation of the critical parameter π_{overlap} .

The meta-analysis yields $\pi_{\text{overlap}} = 0.2452$, indicating that just under a quarter of AOs are concurrent at-risk gamblers. Table 4 summarises this parameter alongside the pooled unconditional prevalences; together, these estimates provide the fixed anchors for the reconciliation model.

3.2 Data Preparation and Raw Distributions

Because Goodwin et al. (2017) fan-out estimates include children, we adjusted them to an adult-equivalent basis. The adult adjustment factor $f_{\text{adult},5+} = P(18+)/P(5+) \approx 0.829$ was derived from Australian Bureau of Statistics demographic ratios. Table 5 presents the adjusted adult-only parameters alongside the original values. As an independent check, we examined household composition data from Hing et al. (2022), which reported that 10.6% of AOs were harmed by a (current or former) partner’s gambling *and* lived in a household with children (i.e., in ‘one parent’ or ‘couple with children’ households). Multiplying this proportion by the average of 1.93 children per Australian household (Australian Institute of Family Studies, 2018) suggests a child-to-adult AO ratio of $\approx 0.205 : 1$, implying an alternative ‘bottom-up’ f_{adult} factor of ≈ 0.830 . This estimate aligns closely with the simple population ratio of 0.829 used throughout the analysis, confirming that our adjustment is robust to alternative calculation methods. Furthermore, sensitivity analysis demonstrates that varying $f_{\text{adult},5+}$ across a plausible range (0.75–0.85) produces negligible shifts in reconciled fan-out estimates.

Notably, these sources differ not just in magnitude but in structure. Self-reports imply a steep severity gradient where harm transmission is negligible among lower-risk gamblers but substantial among PGs. In contrast, AO-proxy reports suggest a flatter distribution with significant harm attributed to LR and MR categories. These conflicting profiles—one concentrated, one diffuse—form the bounds of the reconciliation problem.

To parameterise the right-hand side of the conservation identity (Eq. 5), we require the mean fan-in (μ_{in}). We sourced empirical distributions from two recent Australian surveys, Hing et al. (2022) and Tulloch et al. (2023). To establish a robust baseline, we pooled raw counts reconstructed from the reported proportions and sub-sample sizes ($N = 693$ and $N = 1,131$, respectively), yielding a combined reference distribution of $N = 1819$ (Table 6).

Table 1 Source-study characteristics relevant to AO inclusion and measurement.

Study	AO screening or inclusion wording	AO measurement	Sample and administration
NAT '21 (Hing et al., 2022)	"In the past 12 months, have you been personally affected by another person's gambling?"	Person affecting respondent most; gambling-related harms.	Australian adults; mobile-phone CATI; $N = 15,000$ ($N = 11,560$ asked AO items).
VIC '18 (Rockloff et al., 2020)	"In the last 12 months, have you been personally affected by another person's gambling?"	Person affecting respondent most; Short Harms Scale for CSOs.	Victorian adults; RDD landline/mobile telephone survey; $N = 10,638$.
SA '18 (Woods et al., 2018)	"In the past 12 months, have you been personally affected by another person's gambling?"	Person affecting respondent most; gambling-related harms and relationship.	South Australian adults; CATI landline/mobile survey; $N = 20,017$.
NT '18 (Stevens et al., 2019)	Report wording: negatively affected by "someone else's gambling" in the previous 12 months.	Relationship, gambling type, harm frequency, and help-seeking.	Northern Territory adults; stratified CATI; $N = 4,945$ completed interviews.
NSW '24 (Browne et al., 2024)	"In the past 12 months have you had a close relationship with someone who has gambled?"	GHS-10-AO; one or more harms from another person's gambling.	NSW adults; mobile-only IPND CATI; $N = 10,000$.
ACT (Rockloff et al., 2025)	"In the last 12 months, have you been personally affected by another person's gambling?"	Person affecting respondent most; GHS-10-AO.	ACT adults; mobile-only IPND telephone survey; $N = 10,007$.
Goodwin et al. (Goodwin et al., 2017)	Gamblers: "Has there been a time when your gambling has caused problems in your life, no matter how minor?" AOs: "Have you had a close relationship with a person whose gambling has caused problems in your life, no matter how minor?"	People affected per gambler; gamblers affecting each AO; proxy PGSI for nominated gamblers.	Australian online panel; $N = 5,205$ total (3,076 gamblers; 2,129 AOs).
Tulloch et al. (Tulloch et al., 2023)	Close to someone "who you would say has had problems with their gambling within the past 12 months."	GHS-20-AO; proxy PGSI; number of other people harmed by that person's gambling.	NSW adults recruited through online panel; $N = 1,131$ AOs.
VIC '23 (Suomi et al., 2024)	"As far as you are aware, how many people you have a close relationship with have gambled in the last 12 months?" Follow-up: "In the last 12 months, have you been personally affected by this person's gambling?"	Closest person's perceived severity; GHS-10-AO; additional people affecting respondent.	Victorian adults; mobile-only IPND CATI; $N = 11,000$.

Note. Wording is quoted verbatim where the questionnaire wording was available in the source document; otherwise, the report's own descriptive wording is given. PGSI risk categories were scored consistently across prevalence studies, using 0, 1–2, 3–7, and 8+ for non-problem, low-risk, moderate-risk, and problem gambling respectively. GHS = Gambling Harms Scale; IPND = Integrated Public Number Database; RDD = random-digit dialling.

Table 2 Prevalence of affected other (AO) status and gambler risk levels from Australian jurisdictional surveys.

Study	N	p_{AO}	p_{LR}	p_{MR}	p_{PG}	$p_{\text{G}_{\text{any}}}$	π_{overlap}	Source
NAT '21	11,560	.060	.066	.032	.012	.110	.381	(Hing et al., 2022)
VIC '18	10,638	.061	.067	.024	.007	.098	.221	(Rockloff et al., 2020)
SA '18	20,017	.060	.046	.022	.007	.075	.196	(Woods et al., 2018)
NT '18	5,000	.081	.094	.036	.014	.144	.322	(Stevens et al., 2019)
NSW '24	10,000	.127	.067	.031	.009	.107	.230	(Browne et al., 2024)
ACT '24	10,007	.057	.063	.023	.008	.094	.256	(Rockloff et al., 2025)

Note. $p_{\text{G}_{\text{any}}}$ refers to the prevalence of any at-risk gambler ($p_{\text{LR}} + p_{\text{MR}} + p_{\text{PG}}$). π_{overlap} refers to the proportion of AOs who are also at-risk gamblers, $P(G_{\text{any}} | \text{AO})$. AO status (p_{AO}) was assessed using affected-by-another-gambling screening and harm items whose wording varied across studies; see Table 1. NSW '24 reports notably higher p_{AO} than other jurisdictions; the source of this variation is unclear. Excluding this survey shifts pooled p_{AO} from 0.067 to 0.061, with π_{overlap} essentially unchanged (0.245 to 0.251).

Table 3 Empirical estimates of gambler–AO overlap used in meta-analysis.

Study	π_{overlap} ($P(G_{\text{any}} \text{AO})$)	N_{AO}	Weight (%)
NAT '21	0.381	693	11.2%
VIC '18	0.221	648	14.3%
SA '18	0.196	1,201	28.9%
NT '18	0.322	405	7.0%
NSW '24	0.230	1,270	27.2%
ACT '24	0.256	570	11.4%
Pooled	0.2452	—	100.0%

Note. π_{overlap} ($P(G_{\text{any}} | \text{AO})$) is the proportion of affected others (AOs) who are also at-risk gamblers (LR, MR, or PG). N_{AO} is the effective sample size for the π_{overlap} estimate, calculated as $N \times p_{\text{AO}}$ from Table 2.

3.3 Mixture Model Diagnostics

We evaluated three candidate models against the pooled fan-in counts: a zero-truncated Poisson (ZTP), a zero-truncated negative binomial (ZTNB), and the proposed two-component mixture. The mixture model incorporates the meta-analytic constraint π_{overlap} , assigning a ZTP distribution to the non-overlap component and a ZTNB to the overlap component (subject to $\mu_{\text{overlap}} \geq \mu_{\text{non-overlap}}$). Parameters were estimated by minimising the multinomial negative log-likelihood, treating the 5+ bin as the aggregated tail probability $P(k \geq 5)$.

Single-family models fail to simultaneously capture the distinct mode at $k = 1$ and the heavy tail ($k \geq 5$). The ZTP underestimates the tail, while the ZTNB misfits the relative frequency of dyads ($P(1)/P(2)$). The mixture model resolves this structural

Table 4 Inverse-variance weighted pooled estimates of key model parameters.

Parameter	Notation	Pooled estimate	95% confidence interval
Prev. affected others	p_{AO}	0.0665	(0.0646–0.0684)
Prev. low-risk gamblers	p_{LR}	0.0596	(0.0578–0.0614)
Prev. moderate-risk gamblers	p_{MR}	0.0258	(0.0246–0.0270)
Prev. problem gamblers	p_{PG}	0.0083	(0.0077–0.0090)
Prev. any at-risk gambler	$p_{G_{any}}$	0.0937	(0.0921–0.0966)
AO/gambler overlap	$\pi_{overlap}$	0.2452	(0.2331–0.2572)

Note. The pooled estimate for $p_{G_{any}}$ is the sum of the pooled components (LR, MR, PG) rather than pooled directly, to ensure mathematical consistency (Eq. 1). The 95% confidence interval shown for $p_{G_{any}}$ is from a separate inverse-variance pool of that parameter itself.

Table 5 Goodwin (2017) mean fan-out (μ_{out}) by gambler severity adjusted to adults only.

Severity	AO proxy report		Self report	
	Original	Adult-only	Original	Adult-only
Low Risk Gambler	1.510	1.252	0.030	0.025
Moderate Risk Gambler	3.200	2.653	0.730	0.605
Problem Gambler	5.880	4.875	3.650	3.026

Note. Adult-only means are Goodwin’s AO-inclusive means multiplied by the adjustment factor $f_{adult,5+} \approx 0.829$. This factor is derived from a simple demographic ratio of adults to the population aged 5+ ($P(18+)/P(5+)$).

mismatch (Figure 1), outperforming the best baseline by $\Delta AIC \approx 39$ and accurately recovering the observed shape diagnostics (Table 7). The estimated component means — $\mu_{non-overlap} \approx 1.316$ and $\mu_{overlap} \approx 2.724$ — are carried forward to the reconciliation stage.

Substantively, these results confirm the existence of two distinct subpopulations: a non-gambling majority of AOs ($\approx 75.5\%$) predominantly affected by a single source, and an at-risk minority ($\approx 24.5\%$) subject to multiple harm exposures.

3.4 Solving for the Blending Weight

The fitted mixture model implies a population mean fan-in of $\mu_{in} \approx 1.66$. Substituting this target into the conservation identity (Eq. 5) yields a blending weight of $\alpha = 0.5177$ (95% CI: 0.5097–0.5257). The narrow CI reflects only sampling error; the substantive uncertainty about whether reporting bias is additive or multiplicative is captured by the Value vs Ratio Blend models, which span a 1.8-unit range for μ_{PG} . This near-equal weighting is not imposed by assumption but emerges endogenously from the independent prevalence and fan-in anchors. The result implies that, in terms of aggregate edge volume, the two sources deviate from the network-consistent target by roughly equal magnitudes in opposite directions.

Table 6 Empirical fan-in (k) distribution for affected others (AOs), pooled from counts.

Gamblers (k)	Hing (2022) ($N=695$)	Tulloch (2023) ($N=1131$)	Pooled counts ($N=1819$)	Pooled prop. (%)
1	493	731	1224	67.29%
2	126	262	388	21.33%
3	33	75	108	5.94%
4	10	14	24	1.32%
5+	33	42	75	4.12%
Mean fan-in (μ_{in})	1.509	1.553	1.537	—

Note. $N = 695$ varies slightly from Table 2 ($N = 693$) due to differential missingness on survey items. Counts for Tulloch et al. (2023) were reconstructed from reported proportions and $N = 1131$. The “5+” category is treated as exactly $k = 5$ for calculation of the mean fan-in.

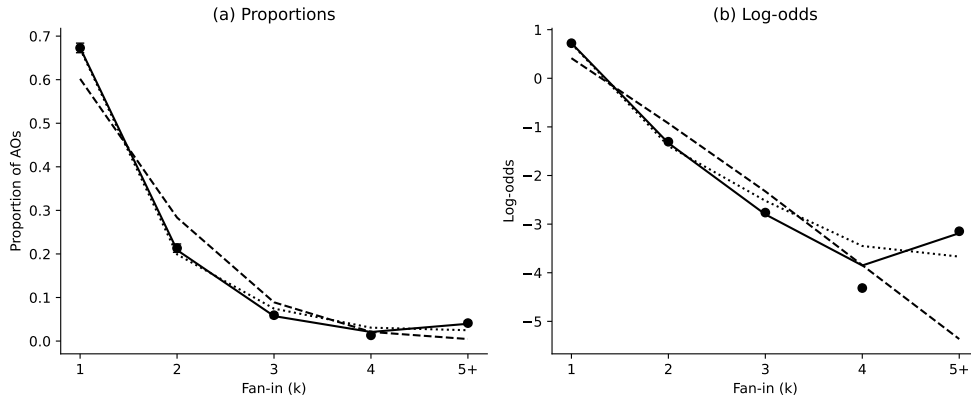


Fig. 1 Observed AO fan-in distribution compared with fitted models. The **Mixture model** (solid line) provides a superior fit to the tail compared to the **ZTP** (dashed line) and **ZTNB** (dotted line) models. **Observed data** are shown as points with ± 1 SE. Panel (a) displays proportions on a linear scale, while Panel (b) uses a log-odds scale to highlight tail divergence.

The Value Blend estimates $\mu_{LR} = 0.62$, $\mu_{MR} = 1.59$, and $\mu_{PG} = 3.92$, while the Ratio Blend yields $\mu_{LR} = 0.25$, $\mu_{MR} = 1.86$, and $\mu_{PG} = 5.73$. This divergence in the high-risk category (3.92 vs 5.73) quantifies structural uncertainty regarding the bias mechanism—additive versus multiplicative—rather than sampling variability. Notably, both solutions impose a substantially steeper severity gradient than the AO-proxy reports privileged by Goodwin et al. (2017), a necessary consequence of reconciling the high fan-in overlap with population prevalence. Conversely, they moderate the extreme sparsity of the self-reports disregarded by that study, recovering material harm contributions from lower-risk groups that self-reporters fail to acknowledge.

Table 7 Model selection and shape diagnostics for pooled AO fan-in.

Model	AIC	$P(1)/P(2)$	$P(k \geq 5)$
Observed	—	3.15	0.0412
ZTP	3736.29	2.12	0.0047
ZTNB	3516.17	3.37	0.0249
Mixture	3477.00	3.22	0.0397

Sensitivity analysis indicates robustness to demographic assumptions. Varying the adult adjustment factor ($f_{\text{adult},5+}$) across a plausible range (0.75–0.85) produces negligible shifts in the reconciled problem-gambler fan-out (μ_{PG} 3.68–3.98 for the Value Blend). This stability is expected, as uniform scalar adjustments to the raw inputs are largely offset by the conservation constraint.

To quantify the adjustment required by reconciliation, we computed the distortion relative to the raw Goodwin data using two metrics: sum of squared errors (SSE) on absolute values and SSE on log-ratios (Table 8). The Ratio Blend minimises shape distortion but incurs significant absolute error. In contrast, the Value Blend proves uniquely robust, outperforming the single-source extremes on *both* metrics. By minimising absolute distortion while simultaneously preserving relative severity gradients better than the raw extremes, the Value Blend offers the most empirically supportable reconciliation of the available evidence.

Table 8 Total adjustment required by each reconciliation scenario under value-based and ratio-based distance metrics. Each blend is optimal under its own bias theory, but both outperform the extreme scenarios under either metric.

Scenario	Value shift (SSE)	Ratio shift (log-ratio SSE)
Value Blend ($\alpha \approx 0.52$)	6.64	13.98
Ratio Blend ($\alpha \approx 0.52$)	16.51	8.92
Affected Other Proxy Only (Scaled, $\alpha = 0$)	9.68	17.74
Gambler Self-Report Only (Scaled, $\alpha = 1$)	53.00	17.58

Note. Value shift is the sum of squared deviations in absolute fan-out values; ratio shift is the sum of squared deviations in log-ratios relative to both raw sources. Values are calculated on an all-persons basis to match the reported fan-out estimates.

While each blended model minimises error under its specific bias theory, the Value Blend proves uniquely robust. As shown in Table 8, it outperforms the single-source extremes on *both* metrics—achieving lower absolute distortion than the AO-only baseline and lower shape distortion than the self-report baseline. This indicates that an additive bias correction yields the most empirically supportable reconciliation, respecting both the magnitude and the relative shape of the original reports better than either of the single-source alternatives.

Table 9 Notation used in the model.

Symbol	Definition
p_{AO}	Prevalence of affected others
p_r	Prevalence of risk category $r \in \{LR, MR, PG\}$
$\mu_{out,r}$	Mean fan-out for risk category r
μ_{in}	Mean fan-in (population average)
$\pi_{overlap}$	$P(G_{any} AO)$, proportion of AOs who gamble
α	Blending weight between self- and AO-proxy reports
f_{adult}	Demographic adjustment factor (0.829)

4 Notation

5 Computational Implementation

All estimation was performed in Python 3.11 using NumPy, SciPy, and pandas. The master reconciliation pipeline (`appendix_a_master_script.py`) implements prevalence pooling via inverse-variance meta-analysis, the closed-form blending weight solution (Eq. 20), both Value and Ratio Blend reconciliation models, extreme-scenario scaling, and all sensitivity analyses. Mixture model fitting (`appendix_b_ao_fanin_models_v3.py`) maximises the multinomial log-likelihood (Eq. 32) using SciPy’s sequential least-squares programming (SLSQP) algorithm, subject to the constraints in Eqs. 34–36, and performs diagnostic comparisons against single-family alternatives. Code and data sufficient to reproduce all reported estimates are archived at [URL will be made available upon manuscript acceptance in accordance to the double-blind requirement].

References

- Australian Institute of Family Studies (2018). *Population and Households* (Tech. Rep.). Australian Institute of Family Studies. Retrieved from <https://aifs.gov.au/research/facts-and-figures/population-and-households>
- Browne, M., Russell, A.M.T., Dellosa, G., Rockloff, M., Hing, N. (2024). *NSW Gambling Survey 2024* (Tech. Rep.). NSW Office of Responsible Gambling. Retrieved from <https://www.gambleaware.nsw.gov.au/resources-and-education/check-out-our-research/published-research/nsw-gambling-survey-2024>
- Goodwin, B.C., Browne, M., Rockloff, M., Rose, J. (2017). A typical problem gambler affects six others. *International Gambling Studies*, 17(2), 276–289, <https://doi.org/10.1080/14459795.2017.1331252>
- Hing, N., Russell, A.M.T., Rockloff, M., Browne, M., Tulloch, C., Rawat, V., . . . Woo, L. (2022). Gambling-related harms to concerned significant others: A national

Australian prevalence study. *Journal of Behavioral Addictions*, 11(2), 361–372, <https://doi.org/10.1556/2006.2022.00045>

- Rockloff, M., Browne, M., Hing, N., Thorne, H., Russell, A., Greer, N., ... Sproston, K. (2020). *Victorian population gambling and health study 2018–2019* (Tech. Rep.). Victorian Responsible Gambling Foundation. Retrieved from <https://responsiblegambling.vic.gov.au/resources/publications/victorian-population-gambling-and-health-study-20182019-759/>
- Rockloff, M., Russell, A.M.T., Browne, M., Hing, N. (2025). *2024 ACT Gambling Survey* (Tech. Rep.). Rockhampton, Queensland: Central Queensland University. Retrieved from https://acquire.cqu.edu.au/articles/report/2024_ACT_Gambling_Survey/29207804
- Stevens, M., Gupta, H., Flack, M. (2019). *Northern Territory Gambling Prevalence and Wellbeing Survey Report, 2018* (Tech. Rep.). Menzies School of Health Research. Retrieved from <https://industry.nt.gov.au/publications/business/publications/gambling-research/2018-nt-gambling-prevalence-and-wellbeing-survey>
- Suomi, A., Kim, J., Biddle, N., Dowling, N.A., Hahn, M., Cowlshaw, S., ... Myers, P. (2024). *Victorian population gambling and health study 2023* (Tech. Rep.). State Government of Victoria. Retrieved from <https://www.justice.vic.gov.au/safer-communities/gambling/victorian-population-gambling-and-health-study-2023>
- Tulloch, C., Browne, M., Hing, N., Rockloff, M., Hilbrecht, M. (2023). How gambling harms others: The influence of relationship-type and closeness on harm, health, and wellbeing. *Journal of Behavioral Addictions*, 12(3), 697–710, <https://doi.org/10.1556/2006.2023.00036>
- Woods, A., Sproston, K., Brook, K., Delfabbro, P., O’Neil, M. (2018). *Gambling prevalence in south australia (2018): final report* (Tech. Rep.). Department of Human Services, South Australia. Retrieved from https://gamblingharmssupport.sa.gov.au/__data/assets/pdf_file/0014/164111/2018-SA-Gambling-Prevalence-Survey-Final-Report-Updated-07.02.19.pdf