Online Resource 2 - Statistical Analysis Plan (SAP v1.1)

ING-01 / T226 | Prospective plan internally frozen on 2 November 2026

Verbatim rendering of the original text file; content not revised post hoc.
The initial offline-frozen status preceded separate launch and funding authorization.

```
ING-01 T2 v0.26 C01 independent replication - prospective SAP v1.1
2026-11-02. OFFLINE FROZEN CANDIDATE; NO LIVE OR FINANCIAL AUTHORIZATION.

1. Population and estimand
80 newly E0-qualified five-agent parents, at most 130 new PRE starts, ordered
T226R01_G001..G130. Stop at 80 qualifications or 130 starts, or the prespecified
financial/technical/time stop. No replacement of qualified technical failures.
The parent is the independent unit. No pooling with R01/E01 or Phase IV.
Target: B+ minus B0 Q3(W12) probability under this fixed T2 task-family mixture
in E0-qualified parents. The whole trigger-and-notice policy is the treatment;
do not condition on B activation, notice receipt or successful publication.

2. Qualification, arms, endpoints
Historical E0/E1/Y code is byte-preserved. PRE <=8 group cycles, two valid PRE
tasks and foreign A information, no PRE B use; empirical E0 also requires valid
semantic audit, neutral manifest, LIVE_LLM and a bound method authorization.
All qualified parents fork once into A, B0, B+, C, each W12=12 group cycles.
A: primary available; B0: primary blocked, pull alternative; B+: same plus the
unchanged neutral notice after first successful agent-initiated B-write; C:
both channels blocked. D is deterministic offline positive control only.
Q3(W12) primary: all three authenticated POST core sources in an eligible
agent's actual submitted model context, confirmed by receipt and SETTLED ledger.
Secondary: Y(W12), historical B0-E1, separately named supported B+ recovery,
Q3/Y(W6), publication/activation/foreign uptake stages. W6 is nested in W12.
Preserve failed A results. Any valid C Q3/Y is an integrity stop, not a discovery.

3. Task contract and allocation
340 new T1 temporal-order instances: 130 ALLOW and 130 HOLD PRE tasks, plus 20
each ALLOW/HOLD/REVIEW/BLOCK POST tasks. All three core sources are necessary
for the full plan; a counterworld need not change the disposition alone.
Historical 60 full records, 41 forbidden factor states and all 16 illustrative
feasibility tasks are excluded as specified in the bank audit.
No claim of 340 novel operative decision problems: there are 33 T2 operative
signatures in the eligible grammar (15 ALLOW,15 HOLD,2 REVIEW,1 BLOCK).
This limitation restricts generalization, not the number of independent parents.
The remaining 20 BLOCK geometries are exhausted; no replacement bank is allowed.
POST assignment is by qualification ordinal, not candidate number. Each four
consecutive ordinals have one of each class. Roles/actor positions balance over
20 qualified parents; A/C positions are balanced overall and by class.
Each parent has a separately drawn fair B-slot coin. B0/B+ balance is NOT forced.
Task/balance seed and complete 130/80 allocation lists are fixed before live.
B0/B+ coins use a separate single 80-bit OS-entropy draw, not the deterministic
task seed; all bits are retained, with no reroll or outcome-based selection.
Within a parent, POST task, actor order, private histories and model contract
are identical across A/B0/B+/C except the specified channel/notice intervention.

4. Primary inference
Pair=(Q3_B0,Q3_Bplus). n01 benefits B+, n10 benefits B0; estimate=(n01-n10)/N_E0.
One two-sided exact paired McNemar/sign test, alpha=.05:
p=min(1,2*BinomialCDF(min(n01,n10),n01+n10,.5)); D=0 gives p=1.
Positive replication requires positive estimate and p<.05. No effect/futility
interims, no optional sample extension, no switch to one-sided testing.
Confirmatory positive replication additionally requires all 80 qualifications.
With N_E0<80, inference is descriptive/sensitivity only, even for small nominal
```

Online Resource 2 - Statistical Analysis Plan (SAP v1.1)

ING-01 / T226 | Prospective plan internally frozen on 2 November 2026

Verbatim rendering of the original text file; content not revised post hoc.
The initial offline-frozen status preceded separate launch and funding authorization.

p: a cost/technical stop can depend on outcomes and does not establish exact
fixed-N inference at a data-dependent sample size. Do not silently re-power it.
The independent B-slot coin justifies a sharp-null label-randomization test
under no cross-branch interference and invariant slot outcomes under that null.
It does not make the test exact for every heterogeneous average-null model.
For marginal-equality inference the conditional exchangeable-pair model applies.
Confidence: separate 97.5% CP bounds for p01,p10 -> [L01-U10,U01-L10], clipped
to [-1,1], conservative >=95% coverage under the independent IID pair model.
This is not necessarily test-compatible. Also report Hoeffding bounds for
independent heterogeneous parent differences: estimate +/-sqrt(2*ln(40)/N).
Fixed-class rates, A/C rates and marginal binomial intervals are descriptive;
no class-specific success claim or multiplicity-adjusted secondary confirmation.

5. Missingness and scope
PRE failures stay in N_all/qualification flow, never as POST zeros. Every
qualified parent stays in N_E0, even if one or all branches are missing.
A receipt-confirmed Q3=1 before a technical stop may survive if the entire
supporting signed/input/receipt prefix is valid; Y is separately host-validated.
Without such a positive, an incomplete W12 is UNKNOWN/CENSORED, never zero.
Invalid evidence is UNKNOWN. Unstarted branches are NOT_STARTED/UNKNOWN.
Enumerate attainable 0/1 completions of missing pairs via discordance counts:
report sharp effect limits, maximum exact p, union of IID-pair intervals, and
union of heterogeneous Hoeffding intervals. Primary positive claim only if all
completions have positive effect and p<.05. Complete-case is supplementary.
No completed-pair exclusion to improve apparent effect or power.

6. Stops, cost, identity and duration
Only new namespace T226_C01_INDEPENDENT_REPLICATION_R01; permanent exclusive
claim before first dispatch; no automatic retry, replay or resume.
At most 24,400 generations, 292,800,000 accepted input tokens and 49,971,200
output tokens; <=12,000 input and <=2,048 output per accepted response.
Model gpt-6.1-sol, medium/default, unchanged local custom tools and prompts.
Counting disabled; cache explicit/30m with no breakpoints; no assumed discount.
450 USD is a proposed nominal ledger cap, NOT APPROVED. Per-parent admission
reserves a complete parent plus one exceptional-response envelope. Stop before
PRE if insufficient; first uncertain outcome stops globally and retains its hold.
Maximum proposed duration 72h; approval validity separately bound. Stop on
UNKNOWN provider result, usage/price/identity drift, failed integrity/isolation,
C success, request/token/time limit or completion. No automatic cap increase.
Alias immutable version not verified; prior acceptance does not authorize this
new series. Document returned model/tier, request IDs, time blocks and any stops.

7. Power and interpretation (reused, not recalculated)
N=80: 84.4% power for delta .25 with p10=.15; 91.4% with p10=.10 under the prior
paired planning model. For delta .15:45.5-73.7% across examined reverse rates.
M=130 completion:98.5% at qualification p=.70,39.6% at p=.60,0.53% at p=.50,
before technical stops. The 62.5pp exploratory pilot result is not the sole
effect assumption. This is an internal independent-parent replication under
one model alias and fixed grammar, not an external-lab/model replication or
proof of spontaneous institutions, universal recovery or common knowledge.

8. Reporting and closure
Report N_all,N_E0,N_complete, all four arm denominators, pair table, W6/W12,
unknown/censored/unstarted counts, class and slot allocations, ledger actuals,

Online Resource 2 - Statistical Analysis Plan (SAP v1.1)

ING-01 / T226 | Prospective plan internally frozen on 2 November 2026

Verbatim rendering of the original text file; content not revised post hoc.
The initial offline-frozen status preceded separate launch and funding authorization.

```
unresolved holds and provider-usage costs separately. Seal raw traces, model
inputs/receipts, signed events and manifest/approval. All additions versioned;
R01/E01 and Phase-IV/P01/P02/P03 remain unchanged. A failed/partial replication
does not permit new sampling without a new prospective study decision.
```