Appendix A

Bayesian Workflow Demonstration and Sensitivity Analysis

(This demonstrates the methods described in Section 6 of the main paper)

DISCLAMER: This appendix presents **purely illustrative Bayesian calculations** using hypothetical data and simplified assumptions. It is intended solely to demonstrate the Bayesian Assessment of Research Causality (BARC) methodological workflow and its sensitivity to key assumptions. **This is not a risk assessment of micro(nano)plastics (MNPs) and hepatocellular carcinoma (HCC).** All parameters, Bayes factors, and priors are chosen for pedagogical clarity and computational demonstration. Real applications would require empirical data, validated priors, and more sophisticated modeling.

A.1 Illustrative Study Setup: Demonstrating Bayesian Updating

To demonstrate how the Bayesian Assessment of Research Causality (BARC) framework operates—particularly in regulatory contexts requiring high certainty—we apply deliberately skeptical priors in this **illustrative example**. This shows how evidence updates beliefs even from conservative starting points.

Illustrative Inputs:

- **Prior probability P(H):** Modeled as Beta(10, 990)
    1. Mean = 0.01 (1%)
    2. Represents regulatory skepticism: "Assume no effect unless strong evidence"
    3. 95% interval: [0.005, 0.017]

Contextual Note on Prior Selection: The skeptical prior Beta(10,990) used in this appendix contrasts with the informed mechanistic prior Beta(85,15) discussed in Section 6 of the main text. This deliberate contrast serves two purposes: (1) to demonstrate Bayesian updating under the conservative assumptions typical of regulatory risk assessment, and (2) to illustrate how the BARC framework accommodates different philosophical starting points through explicit prior specification. Real applications would justify priors through systematic evidence review and expert elicitation.

- **Evidence modeled as Bayes factors (BFs)** [1]**:**

  We use LogNormal distributions to represent uncertainty in evidence strength:

  4. **Mechanistic evidence** (ROS induction in hepatocytes):
     - BF_mech ~ LogNormal(mean=0, sd=0.5)
     - **Note:** Mean log(BF)=0 corresponds to median BF=1 (neutral evidence)
  5. **Human biomonitoring** (hypothetical: 92% vs 67% detection):
     - BF_human ~ LogNormal(mean=ln(3), sd=0.2)
     - **Note:** Represents moderate evidence (median BF=3)
  6. **Animal models** (hypothetical: 3.2-fold increase):
     - BF_animal ~ LogNormal(mean=ln(2), sd=0.2)
     - **Note:** Represents weak evidence (median BF=2)

Derivation Note: The illustrative BFs used here (BF_mech = 1, BF_human = 3, BF_animal = 2) correspond approximately to the following hypothetical study results using the conversion methods in Section A.2:

- **BF_mech = 1:** Represents neutral evidence, e.g., p ≈ 0.37 or OR ≈ 1.1
- **BF_human = 3:** Represents moderate evidence, e.g., OR = 2.0 with 95% CI [1.2, 3.3] → BF ≈ 3.5
- **BF_animal = 2:** Represents weak evidence, e.g., RR = 1.8 with p ≈ 0.1 → BF ≈ 2.0

These round numbers were chosen for **pedagogical clarity** in this workflow demonstration. Real applications would use BFs calculated from actual study data via the methods in Section A.2 or full Bayesian modeling.

Workflow Formulas:

prior_odds = P(H) / (1 - P(H))

BF_total = BF_mech × BF_human × BF_animal

posterior_odds = prior_odds × BF_total

posterior_probability = posterior_odds / (1 + posterior_odds)

Point Calculation Example:

Step 1: prior_odds = 0.01 / (1 - 0.01) = 0.010101

Step 2: BF_total = 1 × 3 × 2 = 6 # Using median values

Step 3: posterior_odds = 0.010101 × 6 = 0.060606

Step 4: posterior_probability = 0.060606 / (1.060606) = 0.0571 (5.71%)

Interpretation: Under these illustrative assumptions, evidence updates probability from 1% to 5.71%. The point calculation (5.71%) uses median Bayes factors, while the simulation mean (6.47%) accounts for the full uncertainty distributions.

A.2 Deriving Bayes Factors: Illustrative Approaches

Important: The formulas below are **simplified approximations** for pedagogical demonstration. Real applications require full Bayesian modeling with proper likelihood specifications.

A. From p-values (Illustrative Approximation):

For studies reporting p-values, the Bayes factor can be approximated using bounds [2]:

$$BF_{10} \geq -\frac{1}{e \cdot p \cdot ln\,(p)}$$

Example: p = 0.03 → BF ≥ 2.7

Interpretation: Moderate evidence against the null.

B. From Odds Ratios (Illustrative Approximation):

For case-control studies with OR and 95% CI [L, U]:

$$SE = \frac{ln(U) - ln\,(L)}{3.92} \qquad BF \approx exp\left(\frac{(ln\,(OR))^2}{2 \cdot SE^2}\right)$$

Example: OR = 2.0, 95% CI [1.2, 3.3] → BF ≈ 3.5

Table A.1: Approximate Mapping of Evidence Strengths to Bayes Factor

(Illustrative Only):

Evidence Strength	Approx. BF	Approx. p-value	Approx. OR Range
Minimal	1-1.5	0.3-0.5	1.1-1.3
Weak	1.5-3	0.1-0.3	1.4-1.7
Moderate	3-10	0.01-0.1	1.8-2.2
Strong	10-30	0.001-0.01	2.3-3.0

Caution: These mappings are oversimplified and context-dependent. They serve only to illustrate how different study outputs might be translated into Bayes factors.

A.3 Sensitivity Analysis via Monte Carlo Simulation

Method: We perform 10,000 Monte Carlo draws to propagate uncertainty through our illustrative Bayesian updating [3] (Python code available in section A.8.1, and also in at : https://github.com/BARC-commits/BARC_Bayesian-Assessment-of-Research-Causality)

Setup:

- Iterations: 10,000
- Random seed: 12345 (for reproducibility)
- Prior: P(H) ~ Beta(10, 990)
- BFs sampled from LogNormal distributions as specified above

[ILLUSTRATIVE EXAMPLE] Table A.2: Illustrative results from hypothetical scenario.

Illustrative results summary.

Parameter	Median	95% Interval	Interpretation
Prior P(H)	1.0% (mean)	[0.5%, 1.7%]	Initial skepticism
BF_mech	1.0 (median)	[0.4, 2.7]	Neutral evidence
BF_human	3.0 (median)	[2.0, 4.5]	Moderate evidence
BF_animal	2.0 (median)	[1.4, 2.9]	Weak evidence
Posterior P(H) (illustrative)	**6.47% (mean)**	**[1.49%, 17.26%]**	**Updated belief**

Interpretation: Even with conservative priors and accounting for uncertainty, the illustrative evidence suggests a meaningful update in probability (hypothetical 6.5-fold multiplier in this demonstration from 1.0% to 6.47%). The wide credible interval (1.49-17.26%) reflects substantial remaining uncertainty.

Mean for probabilities; median for Bayes factors (due to LogNormal skewness). Simulation mean posterior: 6.47%; theoretical Beta mean prior: 1.0% (simulated: 0.995%).

A.4 Sensitivity to Prior Specification

Table A.3 demonstrates how different starting beliefs affect conclusions (Python code available in section A.8.2, and also in at : https://github.com/BARC-commits/BARC_Bayesian-Assessment-of-Research-Causality).

[ILLUSTRATIVE EXAMPLE] Table A3: Hypothetical example showing how priors affect conclusions. **Sensitivity of HCC hypothetical example to prior specification**. Illustrative interpretation: This hypothetical example shows how evidence updates probability. Skeptical priors yield a posterior of 6.47%. Mechanistic-informed priors (chosen for illustration) yield >85%.

Prior type	Prior Parameters	Prior Mean	Posterior Mean	95% Credible Interval
Conservative (Regulatory)	Beta(10,990)	1.0%	6.47% (illustrative)	(1.49%,17.26%)
Informed (Mechanistic)	Beta(85,15)	85.0%	96.68% (illustrative)	(90.91%,99.19%)
Neutral	Beta(50,50)	50.0%	84.14% (illustrative)	(64.02%,95.15%)
Non-informative	Beta(1,1)	50.0%	76.26% (illustrative)	(12.10%,99.61%)

Illustrative insight: In this example, updating direction is consistent across priors (probability increases), but the magnitude depends heavily on starting beliefs. This illustrates why transparent prior specification is crucial.

A.5 Accounting for Evidence Stream Dependence

Important Realism: Evidence streams (mechanistic, animal, human) are rarely independent—they share biological pathways and methodological limitations. Ignoring this dependence overstates combined evidence strength.

Correlation-Adjusted Bayes Factor Calculation:

Let BF_1, BF_2, BF_3 be Bayes factors for three evidence streams with average correlation ρ.

The correction factor is:

$$C(\rho, BF) = 1 + \rho \cdot [(BF_1 \cdot BF_2 - 1) + (BF_1 \cdot BF_3 - 1) + (BF_2 \cdot BF_3 - 1)]$$

Adjusted total BF:

$$BF_{total,corr} = \frac{BF_1 \times BF_2 \times BF_3}{C(\rho, BF)}$$

Illustrative Example:

With BF = [1, 3, 2] and ρ = 0.5:

- C = 1 + 0.5 × [0.73 + 0.41 + 1.45] = 2.30
- BF_total, corr = 6 / 2.30 = 2.61
- Posterior probability = 2.57% (vs. 5.71% under independence)

Sensitivity to Correlation Level:

[ILLUSTRATIVE EXAMPLE] Table A4: Illustrative results from hypothetical scenario. **Effect of varying correlation assumptions (**ρ= 0 to 0.9 between evidence streams) and posterior probabilities (Python code available in section A.8.3, and also in at : https://github.com/BARC-commits/BARC_Bayesian-Assessment-of-Research-Causality).

Correlation (ρ)	Correction Factor (C)	Adjusted total BF	Posterior Probability	Reduction from Independence
0	1.00	6.00	5.71% (illustrative)	0.0%
0.3	1.78	3.37	3.30% (illustrative)	42.3%
0.5	2.30	2.61	2.57% (illustrative)	55.0%
0.7	2.82	2.13	2.11% (illustrative)	63.1%
0.9	3.34	1.80	1.78% (illustrative)	68.8%

Demonstration: When correlation is modeled, it substantially affects results. Moderate correlation (ρ=0.5) reduces posterior probability by ~55% compared to independence assumptions. **This demonstrates why modeling dependence is essential for realistic evidence synthesis.**

A.6 Discussion of Implications for Causal Assessment

This illustrative analysis demonstrates several **key principles** for Bayesian causal assessment [4,5]:

1. **Bayesian updating works** even from skeptical starting points
2. **Uncertainty propagation** is crucial—point estimates mask variability
3. **Evidence dependence** substantially affects conclusions
4. **Prior sensitivity** must be examined transparently

However, critical limitations of this illustrative example include:

- Hypothetical data and simplified distributions
- Oversimplified Bayes factor conversions
- Arbitrary correlation assumptions
- Computational simplicity over realism [6]

In real applications, BARC would require:

- Empirical data with proper likelihood models
- Expert-elicited priors and correlation estimates
- Full hierarchical Bayesian implementation (see Appendix B)
- Validation against known causal relationships

A.7 Note on Regulatory Interpretation

Important: The posterior probabilities calculated here (e.g., 5.71%, 6.5%) are **illustrative numerical outputs** from a pedagogical example. They should **not** be interpreted as actual risk estimates for MNPs and HCC.

If Bayesian posterior probabilities were to be used for regulatory decisions (a speculative future application), they would require:

1. **Empirical calibration** against known carcinogens
2. **Expert consensus** on decision thresholds
3. **Validation** across multiple case studies
4. **Transparent documentation** of all assumptions

The framework demonstrates **how** Bayesian reasoning could inform decisions, not what those decisions should be based on this illustrative example.

A.8 Computational Implementation

The Python code below reproduces these calculations. It is provided for transparency and reproducibility of this **illustrative example** [7].

A.8.1 Python code for illustrative Bayesian calculations

```
"""
FOR DEMONSTRATION PURPOSES ONLY
Illustrative implementation - not for production use
All parameters are hypothetical
"""

# BARC Appendix A: Illustrative Bayesian Calculations
# FOR DEMONSTRATION PURPOSES ONLY

import numpy as np
import scipy.stats as stats

def run_illustrative_simulation(n_iter=10000, seed=12345):
    """
    Run Monte Carlo simulation for illustrative Bayesian updating.
    Returns: prior_probs, posterior_probs
    """
    np.random.seed(seed)

    # Sample from skeptical prior
    prior_probs = stats.beta.rvs(10, 990, size=n_iter)
```

```
    # Sample Bayes factors (illustrative distributions)
    bf_mech = stats.lognorm.rvs(s=0.5, scale=np.exp(0), size=n_iter)
    bf_human = stats.lognorm.rvs(s=0.2, scale=np.exp(np.log(3)), size=n_iter)
    bf_animal = stats.lognorm.rvs(s=0.2, scale=np.exp(np.log(2)), size=n_iter)

    # Calculate posterior probabilities
    posterior_probs = []
    for i in range(n_iter):
        prior_odds = prior_probs[i] / (1 - prior_probs[i])
        bf_total = bf_mech[i] * bf_human[i] * bf_animal[i]
        post_odds = prior_odds * bf_total
        post_prob = post_odds / (1 + post_odds)
        posterior_probs.append(post_prob)

    return np.array(prior_probs), np.array(posterior_probs)

# Example usage
if __name__ == "__main__":
    print("BARC Appendix A: Illustrative Calculations")
    print("=" * 60)
    print("DISCLAIMER: This is a demonstration using hypothetical data.")
    print("=" * 60)

    prior, posterior = run_illustrative_simulation()

    print(f"Prior P(H): {np.mean(prior):.3%} "
          f"[{np.percentile(prior, 2.5):.3%}, {np.percentile(prior, 97.5):.3%}]")
    print(f"Posterior P(H): {np.mean(posterior):.3%} "
          f"[{np.percentile(posterior, 2.5):.3%}, {np.percentile(posterior, 97.5):.3%}]")
```

```
    print(f"Illustrative fold increase: {np.mean(posterior)/np.mean(prior):.1f}x")
```

Output example:

```
BARC Appendix A: Illustrative Calculations
============================================================
DISCLAIMER: This is a demonstration using hypothetical data.
============================================================
Prior P(H): 0.995% [0.479%, 1.711%]
Posterior P(H): 6.470% [1.487%, 17.264%]

Illustrative fold increase: 6.5x
```

* None of these numerical outputs constitute evidence about MNPs and cancer. They demonstrate methodological workflow only.

A.8.2 Python code for sensitivity analysis

```
"""
FOR DEMONSTRATION PURPOSES ONLY
Illustrative implementation - not for production use
All parameters are hypothetical
"""
# BARC_Prior_Sensitivity_Table.py
"""
Generates Table A3: Sensitivity of HCC Hypothetical Example to Prior Specification.
This script runs the Monte Carlo simulation for four different prior distributions
and compiles the results into a comprehensive table.
"""

import numpy as np
import scipy.stats as stats
import pandas as pd

def run_simulation_for_prior(alpha, beta, n_iter=10_000, seed=12345):
    """
    Runs the Monte Carlo simulation for a given Beta prior.

    Parameters:
    alpha, beta (float): Parameters for the Beta(alpha, beta) prior.
    n_iter (int): Number of Monte Carlo draws.
    seed (int): Random seed for reproducibility.

    Returns:
    tuple: (prior_mean, posterior_mean, ci_lower, ci_upper) as raw floats
    """
    np.random.seed(seed)

    # Sample from the specified Beta prior
    prior_samples = stats.beta.rvs(alpha, beta, size=n_iter)

    # Sample Bayes factors (same evidence likelihood for all priors)
    bf_mech   = stats.lognorm.rvs(s=0.5, scale=np.exp(0), size=n_iter)
    bf_human  = stats.lognorm.rvs(s=0.2, scale=np.exp(np.log(3)), size=n_iter)
```

```
    bf_animal = stats.lognorm.rvs(s=0.2, scale=np.exp(np.log(2)), size=n_iter)

    # Compute posterior probabilities
    posterior_probs = []
    for i in range(n_iter):
        prior_odds = prior_samples[i] / (1 - prior_samples[i])
        bf_total = bf_mech[i] * bf_human[i] * bf_animal[i]
        post_odds = prior_odds * bf_total
        post_prob = post_odds / (1 + post_odds)
        posterior_probs.append(post_prob)

    posterior_probs = np.array(posterior_probs)

    # Calculate summary statistics (as raw floats, not percentages)
    prior_mean = alpha / (alpha + beta)
    posterior_mean = np.mean(posterior_probs)
    ci_lower = np.percentile(posterior_probs, 2.5)
    ci_upper = np.percentile(posterior_probs, 97.5)

    return prior_mean, posterior_mean, ci_lower, ci_upper

# Main execution
print("GENERATING TABLE A3: PRIOR SENSITIVITY ANALYSIS")
print("=" * 85)

# Define the four priors to test
priors_to_test = [
    {'name': 'Conservative (Regulatory)', 'alpha': 10, 'beta': 990},
    {'name': 'Informed (Mechanistic)', 'alpha': 85, 'beta': 15},
    {'name': 'Neutral', 'alpha': 50, 'beta': 50},
    {'name': 'Non-informative', 'alpha': 1, 'beta': 1}
]

# Run simulation for each prior and collect results
results = []
for prior in priors_to_test:
    print(f"Running simulation for {prior['name']} prior...")

    prior_mean, post_mean, ci_low, ci_high = run_simulation_for_prior(
```

```
        prior['alpha'],
        prior['beta']
    )

    # Store all values as floats, format only for display
    results.append({
        'Prior type': prior['name'],
        'Prior Parameters': f"Beta({prior['alpha']},{prior['beta']})",
        'Prior Mean': prior_mean,  # Store as float
        'Posterior Mean': post_mean,  # Store as float
        '95% CI Lower': ci_low,  # Store as float
        '95% CI Upper': ci_high  # Store as float
    })

# Create DataFrame
df = pd.DataFrame(results)

# Format for display (without multiplying by 100 incorrectly)
display_df = df.copy()
display_df['Prior Mean'] = df['Prior Mean'].apply(lambda x: f"{x:.1%}")
# Append "(illustrative)" to Posterior Mean outputs
display_df['Posterior Mean'] = df['Posterior Mean'].apply(lambda x: f"{x:.2%}(illustrative)")
display_df['95% Credible Interval'] = df.apply(
    lambda row: f"({row['95% CI Lower']:.3%}, {row['95% CI Upper']:.3%})",
    axis=1
)

# Select only columns to display
final_display = display_df[['Prior type', 'Prior Parameters', 'Prior Mean',
                            'Posterior Mean', '95% Credible Interval']]

print("\n" + "=" * 85)
print("TABLE A3: Sensitivity of HCC Hypothetical Example to Prior Specification")
print("=" * 85)
print(final_display.to_string(index=False))
print("=" * 85)

# Save detailed results to CSV
df.to_csv('BARC_Table_A3_Detailed_Results.csv', index=False)
```

```
print("\nDetailed results saved to 'BARC_Table_A3_Detailed_Results.csv'")

# Also save formatted version
final_display.to_csv('BARC_Table_A3_Formatted.csv', index=False)
print("Formatted table saved to 'BARC_Table_A3_Formatted.csv'")
```

Output example:

```
GENERATING TABLE A3: PRIOR SENSITIVITY ANALYSIS
================================================================
Running simulation for Conservative (Regulatory) prior...
Running simulation for Informed (Mechanistic) prior...
Running simulation for Neutral prior...
Running simulation for Non-informative prior...

================================================================
TABLE A3: Sensitivity of HCC Hypothetical Example to Prior Specification
================================================================
 Prior type Prior Parameters                    Prior Mean      Posterior Mean 95% Credible Interval
 Conservative (Regulatory) Beta(10,990)         1.0%          6.47%(illustrative)    (1.487%, 17.264%)
 Informed (Mechanistic)      Beta(85,15)        85.0%         96.68%(illustrative)    (90.905%, 99.194%)
 Neutral                       Beta(50,50)      50.0%         84.14%(illustrative)    (64.020%, 95.146%)
 Non-informative               Beta(1,1)        50.0%        76.26%(illustrative)    (12.098%, 99.609%)
================================================================

Detailed results saved to 'BARC_Table_A3_Detailed_Results.csv'
Formatted table saved to 'BARC_Table_A3_Formatted.csv'
```

* None of these numerical outputs constitute evidence about MNPs and cancer. They demonstrate methodological workflow only.

A.8.3 Python code to reproduce sensitivity analysis for correlated evidence streams

```
"""
FOR DEMONSTRATION PURPOSES ONLY
Illustrative implementation - not for production use
All parameters are hypothetical
"""
# BARC_Correlation_Adjustment.py
"""
Code to reproduce the sensitivity analysis for correlated evidence streams.
Calculates the correction factor (C) and adjusted posterior probability
for varying levels of correlation (rho).
Matches the methodology in Appendix A.5.
"""

import numpy as np

def calculate_correction_factor(rho, bayes_factors):
    """
    Calculate the correlation correction factor C(rho, BF).

    Parameters:
    rho (float): Correlation coefficient between evidence streams.
    bayes_factors (list): List of Bayes factors for each stream.

    Returns:
    float: The correction factor C.
    """
    n = len(bayes_factors)
    correction = 1.0
    # Sum over all pairs of evidence streams (i < j)
    for i in range(n):
        for j in range(i + 1, n):
```

```
            bf_i = bayes_factors[i]
            bf_j = bayes_factors[j]
            correction += rho * (np.sqrt(bf_i * bf_j) - 1)
    return correction

def adjusted_posterior_probability(prior_prob, bayes_factors, rho):
    """
    Calculate the correlation-adjusted posterior probability.

    Parameters:
    prior_prob (float): Prior probability P(H).
    bayes_factors (list): List of Bayes factors.
    rho (float): Correlation coefficient.

    Returns:
    float: Adjusted posterior probability.
    """
    prior_odds = prior_prob / (1 - prior_prob)
    bf_total_independent = np.prod(bayes_factors) # BF under independence

    if rho == 0:
        correction_factor = 1.0
    else:
        correction_factor = calculate_correction_factor(rho, bayes_factors)

    bf_total_adjusted = bf_total_independent / correction_factor
    post_odds_adjusted = prior_odds * bf_total_adjusted
    post_prob_adjusted = post_odds_adjusted / (1 + post_odds_adjusted)

    return post_prob_adjusted, bf_total_adjusted, correction_factor

# --- Main Analysis, Reproducing Table A3 ---
print("CORRELATION SENSITIVITY ANALYSIS")
print("=" * 70)

# Define baseline parameters (USE YOUR UPDATED BASELINE HERE)
prior_prob = 0.01  # P(H) = 1.0%
```

```
bayes_factors = [1.0, 3.0, 2.0]  # BF_mech, BF_human, BF_animal

# Calculate baseline posterior under independence (rho=0)
base_post_prob, base_bf, _ = adjusted_posterior_probability(prior_prob, bayes_factors, 0)

print(f"Prior P(H): {prior_prob:.1%}")
print(f"Bayes Factors (mech, human, animal): {bayes_factors}")
print(f"\nPosterior under independence (ρ = 0): {base_post_prob:.3%}")
print()

# Define correlation levels to test
correlation_levels = [0, 0.3, 0.5, 0.7, 0.9]

print("Table A4: Effect of Varying Correlation Assumptions")
print("-" * 70)
print(f"{'Correlation (ρ)':<15} {'Correction Factor (C)':<25} {'Adjusted BF_total':<20} {'Posterior
Probability':<20} {'Reduction':<10}")
print("-" * 70)

for rho in correlation_levels:
    post_prob_adj, bf_adj, C = adjusted_posterior_probability(prior_prob, bayes_factors, rho)
    reduction = 100 * (1 - (post_prob_adj / base_post_prob))

    print(f"{rho:<15.1f} {C:<25.3f} {bf_adj:<20.3f} {post_prob_adj:<20.3%} {reduction:<10.1f}%")

print("=" * 70)
```

Output example:

```
CORRELATION SENSITIVITY ANALYSIS
================================================================
Prior P(H): 1.0%
Bayes Factors (mech, human, animal): [1.0, 3.0, 2.0]

Posterior under independence (ρ = 0): 5.714%

Table A4: Effect of Varying Correlation Assumptions
----------------------------------------------------------------
Correlation (ρ) Correction Factor (C)    Adjusted BF_total    Posterior Probability Reduction
----------------------------------------------------------------
0.0             1.000                    6.000                5.714%               0.0    %
0.3             1.779                    3.373                3.295%               42.3   %
0.5             2.298                    2.611                2.570%               55.0   %
0.7             2.817                    2.130                2.106%               63.1   %
0.9             3.336                    1.798                1.784%               68.8   %
================================================================
```

* None of these numerical outputs constitute evidence about MNPs and cancer. They demonstrate methodological workflow only.

KEY TAKEAWAY (DEMONSTRATION ONLY):

- This appendix shows **how** Bayesian updating works
- All numbers are **hypothetical** for demonstration
- The 6.47% posterior is **not** a risk estimate
- Real applications need **real data**

A.9 From Demonstration to Potential Application

This appendix has demonstrated the BARC framework's workflow and methodological principles using a hypothetical scenario. To translate this illustrative example into a practical tool for MNPs-cancer risk assessment, several critical steps would be required.

A.9.1 Data Requirements for Real Application

A real BARC application of MNPs carcinogenicity would require **empirical data** across multiple evidence streams:

Evidence Stream	Minimum Data Requirements	Current Gaps
Human Exposure	- Biomonitoring data from diverse populations - Longitudinal exposure measurements - Tissue-specific accumulation data	- Standardized detection methods - Background exposure levels - Dose-response relationships
Epidemiology	- Large cohort studies with exposure assessment - Case-control studies with exposure biomarkers - Cancer registry linkages	- Long latency considerations - Confounding control - Exposure misclassification correction
Animal Models	- Chronic carcinogenicity bioassays - Multiple species data - Dose-response tumor data	- Relevance of animal models to humans - Extrapolation from high to low doses
Mechanistic	- Pathway activation data - Genotoxicity assays - Omics profiles from exposed systems	- Integration across biological scales - Quantitative pathway modeling

Critical data integration challenges: Temporal alignment of exposure and outcome windows, cross-species extrapolation, and measurement error quantification.

A.9.2 Prior Elicitation Process

Unlike the illustrative priors used in this demonstration, real applications require **systematic, transparent prior specification,** following established guidance on expert knowledge elicitation [8,9].

Structured Expert Elicitation Protocol:

1. **Expert Selection:** Diverse panel including toxicologists, epidemiologists, oncologists, statisticians
2. **Evidence Review:** Systematic review of existing mechanistic and animal data
3. **Quantification:** Use of standardized elicitation methods (e.g., Sheffield protocol, Delphi method) [10].
4. **Documentation:** Transparent reporting of all assumptions and rationales

Example Prior Elicitation Questions:

"Based on current mechanistic evidence, what probability would you assign to MNPs causing oxidative stress sufficient for carcinogenesis?"

"Considering analogies to other particulates (asbestos, PM2.5), how strongly should this inform our prior for MNPs?"

Reporting Standards:

- Prior distributions with justification for parameter choices
- Sensitivity analysis across plausible prior ranges
- Documentation of expert conflicts of interest

A.9.3 Model Validation and Calibration

Before BARC outputs could inform policy, the framework would require **rigorous validation**:

Validation Against Known Carcinogens:

1. **Retrospective Application:** Apply BARC to established carcinogens (tobacco, asbestos)
2. **Calibration Check:** Do posterior probabilities align with IARC classifications?
3. **Predictive Validation:** Test on emerging agents with subsequent outcome data

Performance Metrics:

- **Discrimination:** Ability to distinguish carcinogens from non-carcinogens
- **Calibration:** Agreement between predicted and observed probabilities
- **Reliability:** Consistency across different evidence configurations

Uncertainty Quantification:

- Separate uncertainty from exposure measurement, biological variability, and model structure
- Validate credible interval coverage properties
- Test robustness to missing data scenarios

A.9.4 Implementation Roadmap

Phase 1: Framework Development (1-2 years)

- Develop standardized evidence extraction protocols
- Create prior elicitation templates
- Build computational infrastructure

Phase 2: Pilot Applications (2-3 years)

- Apply to 3-5 well-characterized environmental carcinogens
- Refine methods based on pilot results
- Establish performance benchmarks

Phase 3: MNPs-Specific Assessment (3-5 years)

- Conduct systematic evidence reviews
- Perform expert elicitation for MNPs-specific priors
- Generate initial probabilistic assessments
- Establish monitoring for evidence updates

A.9.5 Quality Assurance Requirements

Technical Standards:

- Open-source code with version control
- Reproducible analysis pipelines
- Comprehensive documentation
- Independent code review

Scientific Oversight:

- Multi-disciplinary advisory board
- Regular methodological review
- Transparent conflict of interest management
- Public comment periods for major assumptions

Regulatory Considerations:

- Alignment with existing evidence evaluation frameworks [11]
- Clear documentation of decision thresholds (if used)
- Protocols for evidence updates and reassessment

A.9.6 Mapping Posterior Probabilities to International Agency of Research on Cancer (IARC) Like Classification Levels

A key interpretative step in risk assessment is translating quantitative posterior probabilities into actionable qualitative classifications. Here we propose a calibration mapping between Bayesian posterior probabilities and IARC style categories, grounded in **Bayesian decision theory**, **empirical calibration of historical IARC evaluations**, and **regulatory risk assessment guidelines**.

Rationale for Calibration:

1. **Bayesian Decision Theoretic Foundation:**

   In Bayesian decision theory, action thresholds should reflect the relative costs of false positive vs. false negative errors [12]. For carcinogen classification—where public health consequences of false negatives typically outweigh those of false positives—the action threshold can be lower than the traditional 95 % frequentist confidence level. A posterior probability **> 50 %** indicates that the hypothesis is more likely than not, warranting at least **Group 2B ("possibly carcinogenic")** attention [13].

2. **Empirical Calibration with Historical IARC Decisions:**

   Retrospective analyses of IARC monographs suggest that agents classified as **Group 2A ("probably carcinogenic")** typically have limited human evidence but sufficient animal or strong mechanistic evidence, corresponding to posterior probabilities in the **70–95 %** range under skeptical priors [14,15].

Conversely, **Group 1 ("carcinogenic")** agents, which require convincing human evidence, align with posteriors **> 95 %**

3. Regulatory guidance on probability-based evidence tiers (US EPA, 2005; EFSA, 2018), which treat probabilities above 70–80% as "likely" or "very likely" causal, warranting preventive action [16,17].

4. **Probabilistic Equivalence to Traditional Standards:**

   A posterior probability of **≈ 89 %** corresponds to a Bayes factor of roughly **8:1** in favor of the hypothesis, which is considered "strong evidence" on Jeffreys' scale. This aligns with the IARC Group 2A requirement of "limited evidence in humans" combined with "sufficient evidence in experimental animals" or "strong mechanistic evidence" [18].

Proposed Calibration Mapping:

Posterior Probability Range	Suggested IARC Like Classification	Interpretation & Recommended Action
< 50 %	Group 3 (Not classifiable)	Evidence does not support carcinogenicity; maintain monitoring.
50–70 %	Group 2B (Possibly carcinogenic)	Evidence is suggestive but limited; warrant further study and precautionary measures.
70–95 %	Group 2A (Probably carcinogenic)	Substantial evidence; strong mechanistic/animal support. Justifies preventive regulation.
> 95 %	Group 1 (Carcinogenic)	Convincing evidence in humans; definitive regulatory action warranted.

Application to the hypothetical HCC Case Study:

In our hypothetical analysis under *informed mechanistic priors* (Table A.3), the illustrative posterior probability reaches **96.68 %**, which would map to **Group 1 (carcinogenic)** under this scheme. Under *conservative regulatory priors*, the illustrative posterior is **6.47%**, which remains in **Group 3 (not classifiable)**—consistent with a skeptical regulatory stance given early evidence.

Limitations and Considerations:

- This mapping is illustrative and should be validated through formal expert elicitation and retrospective analysis of IARC decisions.
- The thresholds are not rigid “bright lines” but continuous gradations; decisionmakers should consider the full posterior distribution and context specific cost benefit tradeoffs.
- As new evidence accumulates, posterior probabilities can be updated, allowing classifications to evolve dynamically—a key advantage over static IARC evaluations.

By calibrating posterior probabilities to established classification frameworks, BARC provides a transparent, quantitative bridge between Bayesian inference and regulatory decision making. This approach makes the strength of evidence explicit, supports consistent communication of risk, and enables adaptive classification as science advances.

A.9.7 Limitations and Challenges

Even with ideal implementation, several challenges would remain:

1. **Evidence Quality:** Garbage in, garbage out principle applies
2. **Correlation Estimation:** Difficult to empirically estimate evidence stream dependencies
3. **Computational Complexity:** Hierarchical models require significant expertise
4. **Communication:** Translating posterior probabilities to policy decisions
5. **Dynamic Evidence:** Framework must adapt to rapidly evolving science

Conclusion: From Demonstration to Decision Support

This illustrative appendix has shown **how** BARC could work. Real application would require:

- **Substantial empirical data** collection and standardization
- **Transparent, structured processes** for prior specification
- **Rigorous validation** against known carcinogens
- **Ongoing refinement** as evidence accumulates

The BARC framework offers a **principled approach** to evidence synthesis under uncertainty, but its transition from methodological demonstration to decision-support tool requires careful, stepwise implementation with appropriate validation at each stage.

The hypothetical numbers in this appendix demonstrate methodology; real numbers for MNPs-cancer assessment await this implementation pathway.

Appendix References:

1. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90(430):773-95.

2. Sellke T, Bayarri MJ, Berger JO. Calibration of p-values for testing precise null hypotheses. Am Stat. 2001;55(1):62-71.

3. Robert CP, Casella G. Monte Carlo statistical methods. New York: Springer; 2004.

4. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med. 1999;130:1005-1013.

5. Gelman A, Vehtari A, Simpson D, et al. Bayesian workflow. arXiv [Preprint]. 2020. Available from: https://doi.org/10.48550/arXiv.2011.01808

6. Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo. arXiv [Preprint]. 2017. Available from: https://arxiv.org/abs/1701.02434

7. Abril-Pla O, Andreani V, Carrool C, et al. PyMC: a modern probabilistic programming language in Python. PeerJ Comput Sci. 2023;9:e1516.

8. EFSA Scientific Committee. Guidance on expert knowledge elicitation in food and feed safety risk assessment. EFSA J. 2014;12:3734.

9. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press; 2007.

10. O'Hagan A, Buck CE, Daneshkhah A, et al. Uncertain judgements: eliciting experts' probabilities. Chichester (UK): Wiley; 2006.

11. US Environmental Protection Agency. Guidance for considering and using open literature toxicity studies to support human health risk assessment. Washington (DC): US EPA; 2014.

12. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90(430):773–95.

13. Jeffreys H. Theory of probability. 3rd ed. London: Oxford University Press; 1961.

14. Baan R, Straif K, Grosse Y, et al. Carcinogenicity of some aromatic amines, organic dyes, and related exposures. Lancet Oncol. 2008 Apr;9(4):322–3.

15. Cogliano VJ, Baan R, Straif K, et al. Preventable exposures associated with human cancers. J Natl Cancer Inst. 2011;103(24):1827–39.

16. US Environmental Protection Agency (EPA). Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. Washington, DC: US EPA; 2005. Available from: https://www.epa.gov/risk/guidelines-carcinogen-risk-assessment

17. EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, et al. Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J. 2017 Aug;15(8):e04971. doi: https://doi.org/10.2903/j.efsa.2017.4971

18. Goodman S, Greenland S. Assessing the unreliability of the medical literature: a response to “Why most published research findings are false.” Johns Hopkins University, Dept. of Biostatistics Working Papers. 2007; Working Paper 135. Available from: https://biostats.bepress.com/jhubiostat/paper135