

Nancy Truong

19/09/2025

# Evidence synthesis: Report

## Problem:

We analyse a dataset set of a collection of randomised case-control studies that measured the effectiveness of a descriptive social-norm intervention ("social") vs control on hotel guests' towel reuse. The dataset contains the number of guests who reused their towels ( `reuse` ) and the total number of guests observed ( `total` ) for each `study` and `group` .

## Reference:

Scheibehenne, B., Jamil, T., & Wagenmakers, E.-J. (2016). Bayesian Evidence Synthesis Can Reconcile Seemingly Inconsistent Results: The Case of Hotel Towel Reuse. *Psychological Science*, 27(7), 1043-1046.

<https://journals.sagepub.com/doi/10.1177/0956797616644081>

## Aim:

1. Formulate and justify a parametric model for testing the effectiveness of the intervention (i.e. `group` ).
2. Estimate the model parameters.
3. Test if the intervention has an effect or not.

## Preparing data

```
In [1]: # Libraries bambi
        #pip install bambi
        import pandas as pd
        import numpy as np
        import bambi as bmb

        print("Import libraries done.")
```

Import libraries done.

## Data wrangling

```

In [2]: data_file = "/Users/mexmex/Documents/3-Math_LU/HT2025/BERN02/Exercise 6/Data/towelD
data = pd.read_csv(data_file, sep=';', encoding='latin1')
count = data.iloc[:, -1] # get the last column with numbers or yes/no

# count has the number of yes and no for control and social norm groups
control_yes = count[::4].to_numpy() # every 4th starting from 0 - control group + y
control_no = count[2::4].to_numpy() # every 4th starting from 2 - control group + n
control_total = np.array([y + n for y, n in zip(control_yes, control_no)])

social_yes = count[1::4].to_numpy() # every 4th starting from 1 - social norm group
social_no = count[3::4].to_numpy() # every 4th starting from 3 - social norm group
social_total = np.array([y + n for y, n in zip(social_yes, social_no)])

study = np.arange(1, len(control_yes)+1) # 7 different studies

control_data = pd.DataFrame({"reuse": control_yes, "total": control_total, "group":
social_data = pd.DataFrame({"reuse": social_yes, "total": social_total, "group": "

combined_data = pd.concat([control_data, social_data], ignore_index=True)

# Convert data types - important for bambi that they are correctly set
# can give errors otherwise
combined_data['reuse'] = combined_data['reuse'].astype(int)
combined_data['total'] = combined_data['total'].astype(int)
combined_data['group'] = combined_data['group'].astype('category')
combined_data['study'] = combined_data['study'].astype('category')

```

```

In [3]: combined_data

```

Out[3]:

	reuse	total	group	study
0	74	211	control	1
1	103	277	control	2
2	77	135	control	3
3	82	187	control	4
4	21	25	control	5
5	123	147	control	6
6	28	30	control	7
7	98	222	social	1
8	587	1318	social	2
9	406	655	social	3
10	278	555	social	4
11	21	24	social	5
12	472	576	social	6
13	101	132	social	7

## 1. Formulate and justify a parametric model for testing the effectiveness of the intervention (i.e. `group`).

Our main effect of interest is `group`, which indicates whether participants are in the control group or the social-norm intervention group. Let  $i = 1, \dots, 7$  index the different studies and  $j = 1, 2$  index the two group within each study. We have

$$y_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}),$$

where  $y_{ij}$  is the number of people who reused their towels in observation of study  $i$  and group  $j$ ,  $n_{ij}$  is the total number of people in that study and group, and  $p_{ij}$  is the probability of that a person in that observation reuses their towel.

To account for variation in baseline reuse probability between studies, we use a hierarchical binomial model with a varying intercept for each study. We model the log-odds of `reuse` as a function of a fixed effect for the group and a random effect for the study:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \text{study}_i + \beta_1 \cdot \text{group}_i,$$

for  $i = 1, \dots, 14$ ; where:

- $\beta_0$  : Overall intercept of log-odds of `reuse` when `group == "control"`.
- $\text{study}_i$  : Random intercept for `study`  $i$ .

- $\beta_1$  : Fixed effect of group social vs control.
- $group_i = 1$  if social, 0 otherwise.

## 2. Estimate parameters using Bayesian inference

```
In [4]: RANDOM_SEED = 48 # for reproducibility
model0 = bmb.Model("p(reuse, total) ~ group + (1|study)", combined_data, family="bi
model0
```

```
Out[4]:      Formula: p(reuse, total) ~ group + (1|study)
          Family: binomial
          Link: p = logit
Observations: 14
          Priors:
target = p
Common-level effects
  Intercept ~ Normal(mu: 0.0, sigma: 1.5)
  group ~ Normal(mu: 0.0, sigma: 1.0)

Group-level effects
  1|study ~ Normal(mu: 0.0, sigma: HalfNormal(sigma: 2.5495))
```

The model specifies that the probability of reuse ( $p$ ) depends on the group (control or social norm) and includes a random intercept for each study to account for variability between studies.

The prior parameters are set to default values in bambi, which are generally weakly informative priors suitable for most applications. For the fixed effects (group), bambi uses a normal prior with mean 0 and standard deviation 1 on the log-odds scale.

For the random effects (study), the intercept for each study is drawn from a normal distribution with mean 0 and standard deviation 1. The standard deviation is drawn from a HalfNormal(2.5495). HalfNormal is used to ensure that the standard deviation is positive, modeling how much studies can vary from the overall intercept. The choice of HalfNormal(2.5495) as the prior for the standard deviation of the random effects is based on the recommendation by Gelman (2006) to use a weakly informative prior that allows for reasonable variability without being too restrictive.

### Model fitting

The model is fitted using the `fit` method, Markov Chain Monte Carlo (MCMC) sampling via Bambino. We set the random seed for reproducibility.

With `tunes=2000`, the first 2000 iterations of each MCMC were used for tuning (warm-up), allowing the sampler to adapt to the posterior distribution.

After tuning, 2000 posterior (`draws=2000`) were used for inference, giving estimates for model parameters.

`target_accept=0.95` make the step size of the sampler adjusted to achieve a high acceptance probability, reducing the risk of divergent transitions and improving sampling stability.

The fitting procedure generates the posterior distribution for all model parameters, including the intercept  $\beta_0$ , the group effect  $\beta_1$  (social-norm intervention), and the study random intercepts  $study_i$ .

```
In [5]: idata_model0 = model0.fit(tune=2000, draws=2000, target_accept=0.95, random_seed=RAN
```

```
Initializing NUTS using jitter+adapt_diag...
Multiprocess sampling (4 chains in 4 jobs)
NUTS: [Intercept, group, 1|study_sigma, 1|study_offset]
Output()
```

```
Sampling 4 chains for 2_000 tune and 2_000 draw iterations (8_000 + 8_000 draws total) took 11 seconds.
There was 1 divergence after tuning. Increase `target_accept` or reparameterize.
```

## 4. Make a summary of the parameter values by the posterior mean and 90% probability interval.

```
In [6]: import arviz as az

az.summary(idata_model0, hdi_prob=0.90)
```

```
Out[6]:
```

	mean	sd	hdi_5%	hdi_95%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_
<b>Intercept</b>	0.427	0.425	-0.285	1.082	0.011	0.008	1599.0	2497.0	
<b>group[social]</b>	0.208	0.078	0.081	0.336	0.001	0.001	4551.0	4247.0	
<b>1 study_sigma</b>	1.114	0.412	0.537	1.679	0.011	0.012	1492.0	2655.0	
<b>1 study[1]</b>	-0.943	0.429	-1.586	-0.202	0.011	0.008	1618.0	2699.0	
<b>1 study[2]</b>	-0.868	0.423	-1.546	-0.175	0.011	0.008	1599.0	2530.0	
<b>1 study[3]</b>	-0.143	0.426	-0.826	0.558	0.011	0.008	1610.0	2498.0	
<b>1 study[4]</b>	-0.638	0.425	-1.282	0.101	0.011	0.008	1629.0	2590.0	
<b>1 study[5]</b>	1.130	0.543	0.238	2.008	0.012	0.008	2173.0	2990.0	
<b>1 study[6]</b>	0.939	0.428	0.264	1.643	0.011	0.008	1625.0	2598.0	
<b>1 study[7]</b>	0.749	0.454	-0.001	1.468	0.011	0.007	1787.0	2840.0	

### Interpretation of the results

1. The posterior mean of the intercept ( $\beta_0$ ) is 0.427 with a 90% credible interval of  $[-0.285, 1.082]$ . This represents the log-odds of towel reuse in the control group

(baseline category) averaged across all studies. The positive value indicates that, on average, there is a tendency towards towel reuse in the control group.

- On the probability scale, this corresponds to a mean probability of towel reuse of approximately 60.5 (calculated as

$$\frac{1}{1 + \exp(-0.427)}$$

), with a 90% credible interval of approximately [42.9%, 74.7%].

2. Group effect for social  $\beta_1$  posterior mean = 0.208 with positive credible interval [0.081, 0.336], which indicates evidence that the social-norm intervention increases towel reuse.

- On the probability scale, adding 0.208 in log-odds increases the mean probability of towel reuse to approximately 65.4 (calculated as

$$\frac{1}{1 + \exp(-(0.427 + 0.208))}$$

), with a 90% credible interval of approximately [62.4%, 68.2%].

3. Study random effects  $study_i$

- $\sigma = 1.114$  shows certain variation between studies in baseline reuse probability.
- Individual study intercepts  $study_i$  shifts the slope up and down for each study. The intercepts differ between studies, justifying the hierarchical (random intercept) model.
  - Study 3, 4 and 7 both have credible intervals that contain 0, suggesting baseline reuse probability in these studies may not differ significantly from the overall average. However, study 7's credible interval barely includes 0, we cannot conclusively say it differs from the overall baseline.
  - Study 5 has the highest positive deviation (1.130), meaning higher baseline reuse probability.
  - Study 1 has the largest negative deviation (-0.943), meaning lower baseline reuse probability.

## 5. referring to one or several parameters within the model. referring to one or several parameters within the model.

Since we are interested in testing whether the social-norm intervention increases towel reuse, our hypotheses are:

$$H_0 : \beta_1 \leq 0 \quad (\text{the intervention has no effect or reduces reuse})$$

$$H_1 : \beta_1 > 0 \quad (\text{the intervention increases reuse})$$

## 6. Test the hypothesis using Bayesian inference.

We test this hypothesis by examining the posterior distribution of  $\beta_1$  from the fitted model.

```
In [7]: posterior_group = idata_model0.posterior["group"].values.flatten()
print("Bayesian p-value: ", np.mean(posterior_group < 0))
# p-value very small, reject H0 meaning that group is statistically significant
```

Bayesian p-value: 0.002875

`posterior_group` contains all posterior samples for the parameter `group`. We compute the proportion of samples that are less than 0. If this proportion is very small, it indicates that most of the posterior sample is above 0 and we reject the null hypothesis  $H_0$  in favor of the alternative hypothesis  $H_1$ .

Bayesian p-value is very small ( $= 0.002875$ ). This means that we can reject the null hypothesis  $H_0$ . We conclude that the group effect is statistically significant, and the social-norm intervention effectively increases towel use.