



MWC DISCUSSION PAPER No.1

ベイズ的インパクト評価による政策判断

— 推定結果をどう読むか —

株式会社メトリクスワークコンサルタンツ

青柳 恵太郎

初稿完成：2026年1月

オンライン公開：2026年6月

要旨

インパクト評価は、政策や事業の採否を判断するための材料を得ることを目的として実施される。しかし、標準的なインパクト評価で提示される点推定値や信頼区間は、介入が採択に値する改善をもたらす可能性(十分性)や、望ましくない結果をもたらす可能性(安全性)について、政策判断に必要な情報を必ずしも直接には与えない。本稿では、まず十分性と安全性という二つの観点から、信頼区間を用いた標準的な結果の読み方を整理する。その上で、ベイズ的アプローチによって得られる事後分布を用いると、成功確率や害リスクといった形で判断に必要な情報を確率として表現できることを示す。さらに、トルコの職業訓練プログラムのRCTと、ケニアの無条件現金給付(UCT)のRCTを事例として取り上げ、十分性と安全性、ならびに悪影響の許容可能性に関する判断が、事後分布によってどのように整理されるのかを具体的に示す。これにより、ベイズ的インパクト評価が、推定結果を政策判断へと結び付けるための有用な視点を提供することを論じる。

1. はじめに

政策決定者は、どのような情報を得ることを期待してインパクト評価を行うのだろうか。検討の俎上にある介入案について、効果の大きさや副作用の有無がすでに概ね見通せているのであれば、追加的に因果効果を推定する必要性は小さい。多くの場合、インパクト評価が求められるのは、介入案の効果に関する実証的な情報が乏しく、採択すべきかどうかを判断するだけの材料が揃っていない局面である。したがってインパクト評価への期待は、介入案の採択に関する判断を支えるための、より確かな材料を得ることに向けられる。

この判断目的に照らすと、少なくとも次の二つの観点に関する情報が不可欠となる。第一は、「その介入は採択に値する改善をどの程度もたらすと考えられるか」という十分性である。第二は、「その介入は望ましくない結果（状況の悪化）をもたらしてしまう恐れはどの程度あるのか」という安全性である。

本稿では、意思決定者の期待（十分性と安全性）に照らして、インパクト評価結果の標準的な提示形式が提供できる情報と、その形式だけでは残る限界を整理する。そして、ベイズ的アプローチがその限界を補うための見通しを与えることを示す。

2. 標準的インパクト評価における介入効果と十分性・安全性の読み方

介入効果は、概念的には「介入があった世界」と「なかった世界」のアウトカムの差（反事実差）として定義される。介入効果自体は個々の受益者単位で定義できる概念であるが、政策決定者が介入案の採否を判断するうえでは、対象集団において介入効果が平均して好ましい方向に発現しているかを問うことには一定の合理性がある。もちろん平均効果だけでは分配や個人リスクは捉えきれないが、政策判断に必要な情報を要約する指標として対象集団の平均介入効果（Average Treatment Effect : ATE）を用いることが実務上広く行われている。そこで本稿においても、ATEに基づいて十分性と安全性を読む枠組みを整理する。

介入効果の集団平均である ATE は $\tau \equiv E[Y(1) - Y(0)]$ と定義される。しかし、この定義に含まれる反事実（潜在アウトカム）は同一個体について同時に観測することができない（因果推論の根本問題（fundamental problem of causal inference）と呼ばれている）。そこで、実際の推定においては、ランダム割付に代表される研究デザイン、あるいは観察データにおける識別仮定にもとづいて介入群と比較群を構成し、両群のアウトカムの平均（または割合）の差を ATE の推定量として用いる（ $\hat{\tau} = \bar{Y}_T - \bar{Y}_C$ ）。さらに、推定値には標本誤差が伴うため、最終的な推定結果としては点推定値に加えて信頼区間を提示することが標準的である。信頼区間は、分析に用いたデータから見て ATE の真値として候補に残る値の範囲を表す。

したがって、注目している値が信頼区間の外にあるなら「その値は候補から外れやすい」、区間の内にあるなら「外しきれない」と読むことができる。ここでは信頼区間を下限 L と上限 U を用いて $[L, U]$ と表す。たとえば 95%信頼区間は、点推定値 $\hat{\tau}$ と標準誤差 $SE(\hat{\tau})$ にもとづき $CI_{0.95} = [\hat{\tau} - 1.96 \times SE(\hat{\tau}), \hat{\tau} + 1.96 \times SE(\hat{\tau})] \equiv [L, U]$ のように表される。

以下では、この推定結果を十分性と安全性の観点から解釈するために、二つの境界を導入する。第一に、十分性の境界として最低限受容可能な ATE の水準を θ とおく。これは、平均的な改善がどの程度あれば採択に値するとみなすかという判断上の基準である。第二に、安全性の境界として κ をおく。平均介入効果が κ を下回ることは、少なくとも平均としては望ましくない変化を意味する。本稿では議論を単純化するため $\kappa = 0$ とし、平均介入効果が 0 を下回る可能性が残るかどうかを安全性の観点として扱う（以後、 κ は 0 として議論する）。

この二つの境界 (θ と 0) により、ATE が取り得る位置は三つの領域に区切られる。すなわち、十分領域 ($ATE \geq \theta$)、十分未満だが非害の領域 ($0 \leq ATE < \theta$)、害領域 ($ATE < 0$) である (図 1)。したがって、推定結果は「信頼区間 $[L, U]$ が 0 と θ に対してどこに位置しているか」という形で解釈できる。推定された信頼区間をこの二つの境界に重ねて読むことで、十分性と安全性に関する状況を整理するというのが、冒頭の期待に対する標準的なインパクト評価の応答形式である。

図 1 十分性と安全性による ATE の領域区分



表 1 は、その読み方を四つの状態として整理したものである。状態 1 は $L \geq \theta$ の場合であり、信頼区間全体が θ 以上に位置するため、「 θ 以上」という十分性を支持できる。同時に $L > 0$ でもあるため、「0 未満」も排除できる。状態 2 は $U < \theta$ の場合であり、信頼区間全体が θ に届かないため、「 θ 以上」は排除でき、十分性は支持できない。一方で、この条件だけでは 0 との関係は確定しない (安全性は未確定である)¹。状態 3 は $L > 0$ かつ $L <$

¹ 状態 2 は 0 との位置関係により、残る論点は異なり得る。たとえば信頼区間が 0 より上にある場合は「効果は小さいが害は示唆されにくい」という含意を持つ一方、0 を含む場合は「害の可能性を排除できない」という含意が残る。さらに、信頼区間が 0 より下にある場合は「平均として悪化が示唆される」という含意を持つ。ただし、いずれの場合も「十分性 (θ 以上) が支持されない」という本状態の結論は変わらないため、この区別にはこれ以上立ち入らない。

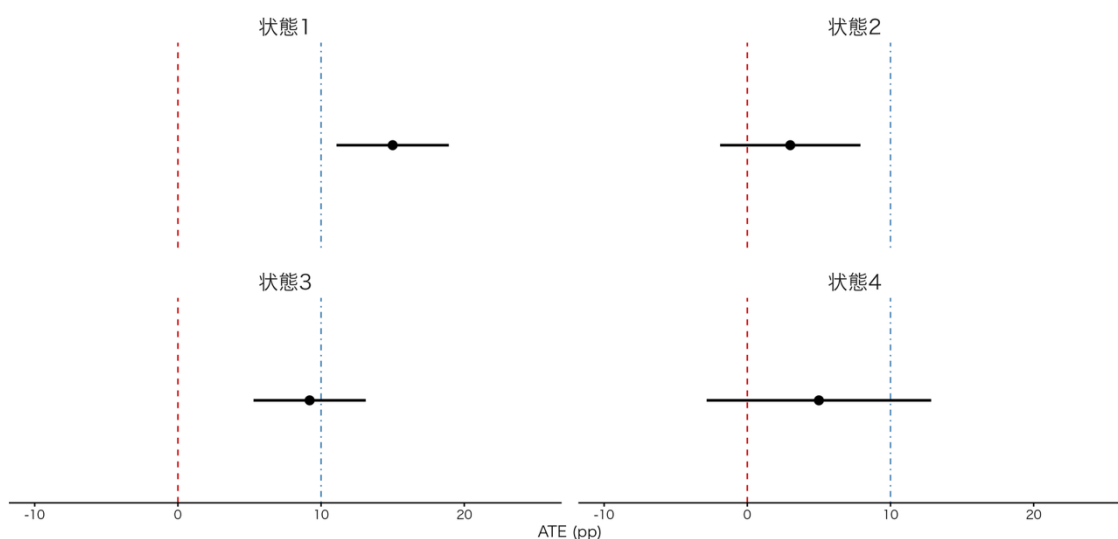
$\theta < U$ の場合であり、「0 未満」は排除できる一方で、 θ は区間内に残るため、「 θ 以上」は未確定として残る。状態 4 は $L \leq 0$ かつ $L < \theta < U$ の場合であり、 0 も θ も区間内に残るため、十分性も安全性も未確定として残る。

表 1 信頼区間の位置による 4 状態の整理

状態	条件 (CIのL・Uと θ の関係)	十分性について (θ 以上の可能性)	安全性について (0 未満の可能性)
1	$L \geq \theta$	支持できる	排除できる
2	$U < \theta$	排除できる	未確定 (可能性は残る)
3	$L > 0$ かつ $L < \theta < U$	未確定 (可能性は残る)	排除できる
4	$L \leq 0$ かつ $L < \theta < U$	未確定 (可能性は残る)	未確定 (害リスクが残る)

図 2 は、表 1 で整理した四つの状態を、仮想の数値例を用いて直感的に示したものである。縦線は 0 (害の境界) と θ (十分性の境界) を表し、点が点推定値、横線が信頼区間である。ここで用いる数値例では、施設分娩の有無のような二値アウトカムを念頭に置き、介入効果を割合 (確率) の差として表す。十分性の閾値を $\theta = 10$ (10 ポイント)、安全性の境界を 0、信頼水準を 95% ($z \approx 1.96$) とする。

図 2 信頼区間の位置による 4 状態 (数値例)



まず、状態 1（十分性が高く安全性も高い）の例として、 $\hat{\tau} = 15$ 、 $SE = 2$ を挙げる。95%信頼区間は $15 \pm 1.96 \times 2 = [11.1, 18.9]$ となり、区間全体が $\theta = 10$ を上回る。したがって十分性が高いと整理できる。さらに区間は 0 も自然に上回るため、平均として悪化しているという見方も排除できる。

次に、状態 2（十分性は示されない）の例として、 $\hat{\tau} = 3$ 、 $SE = 2.5$ を考える。このとき 95%信頼区間は $[-1.9, 7.9]$ となり、区間全体が $\theta = 10$ より下にある。したがって「 θ 以上」は排除され、十分性は支持されない。なお、この例では 0 もまたいでいるため、平均として悪化していないと言い切ることもできないが、いずれにせよ十分性が示されないという点は変わらない。

続いて、状態 3（0 未満は排除されるが十分性は未確定）の例として、 $\hat{\tau} = 9.2$ 、 $SE = 2$ を置く。95%信頼区間は $[5.3, 13.1]$ となり、区間は 0 より上にあるので 0 未満は排除できる。しかし $\theta = 10$ は区間内に残るため、十分性を言い切ることはできない。したがって「平均として悪化している」という懸念は弱められる一方で、「採択に値するほど十分かどうか」は未確定として残る。

最後に、状態 4（十分性も安全性も未確定）の例として、 $\hat{\tau} = 5$ 、 $SE = 4$ を仮定する。95%信頼区間は $[-2.8, 12.8]$ となり、区間が 0 も θ もまたぐ。十分性の可能性も残るが、0 未満も排除できないため、平均として悪化しているという懸念も残る。

以上が、意思決定者の期待（十分性と安全性）に照らして、頻度論的インパクト評価が提供する標準的な状況整理である。この整理を意思決定の観点から読むと、状態 1 では十分性（ θ 以上）と安全性（0 未満の否定）の両方が支持されるため、採択という判断にほぼ迷いが無い。状態 2 では少なくとも十分性（ θ 以上）は成り立たないため、不採択という判断が導かれる（信頼区間が 0 をまたぐかどうかにかかわらず、十分性が満たされないという点は変わらない）。判断が止まりやすいのは状態 3 と状態 4 である。いずれも「 θ 以上」を排除できないため採択の可能性は残る一方で、状態 3 では十分性だけが未確定として残り、状態 4 では十分性に加えて 0 未満の可能性も排除できない。したがって、特に状態 4 では安全性がここで初めて実質的な論点として前面化し、採択と不採択の間で判断が停滞しやすい²。

信頼区間による整理は、「何が支持され、何が排除され、何が未確定として残るか」を明確

² 頻度論においても、信頼区間が境界値（例えば θ ）を含む場合に、判断が直ちに停止するわけではない。推定値と境界値との差を標準誤差で正規化した量（例： $(\hat{\tau} - \theta)/SE$ ）を併用すれば、「境界からの距離」という形で、平均介入効果（ATE）が θ を上回っているかを検討するための情報を得ることができる。さらに、尤度比を用いれば、「ATE がちょうど θ である」という仮定がデータに照らしてどの程度もっともらしいかを、点推定（データに最も整合的な値）と比較して要約できる。もっとも、これらは「 τ が θ 以上である見込み」等を確率として与えるものではない。

にする点では有用である。しかし状態3や状態4では、十分性 ($\tau \geq \theta$) や害 ($\tau < 0$) について、それぞれが成立する確率を明示できないため、意思決定の材料が不足する。信頼区間は境界をまたぐかどうかを示すことはできても、こうした確率を意思決定に使える形で提示しない。次節では、同じ 0 と θ という境界に対して、十分性と安全性をより直接に評価できる形で示すために、ベイズ的アプローチを導入する。

3. ベイズ的インパクト評価の出力：事後分布と境界上の確率

3.1. 事後分布とは何か

前節で見たとおり、信頼区間を 0 と θ に重ねて読むことで、十分性と安全性について「何が支持され、何が排除され、何が未確定として残るか」を整理することはできる。しかし、状態3や状態4のように判断が停滞しやすい局面では、十分性 ($\tau \geq \theta$) や害 ($\tau < 0$) がどの程度の確率で成り立つのかが判断材料として重要になる一方で、信頼区間の提示だけからその確率を読み取ることはできない。ここが、意思決定者の関心（十分性と安全性）と、頻度論の標準出力（点推定と信頼区間）の間に残りやすい摩擦である。

ベイズ的推定では、推定結果を「介入効果はどの値があり得そうか」という形の確率分布として提示する。この確率分布を τ の事後分布と呼ぶ。ここで重要なのは、ベイズ的推定の出力では推定結果が「点」や「区間」ではなく「代表値の周りにどれくらい不確実性が残っているか」という「分布」として提示されるという点である³。前節の信頼区間が、境界に照らして「候補から外れやすい／外しきれない」を整理する道具だとすれば、事後分布は、 θ 以上や 0 未満がどの程度の確率として残るかを数値で示す。すなわち、

- ・ 成功確率（十分性）： $\Pr(\tau \geq \theta \mid data)$
- ・ 害リスク（安全性）： $\Pr(\tau < 0 \mid data)$
- ・ 中間域（非害だが十分未満）： $\Pr(0 \leq \tau < \theta \mid data)$

である。したがって、状態3や状態4のように境界をまたぐ局面でも、議論は「未確定」という整理で止まるのではなく、成功確率と害リスクを材料に前へ進めやすくなる。

³ ベイズでも、事後平均や信用区間（事後分布の一定割合を含む区間）といった要約統計は用いられる。本稿では区間提示自体よりも、 0 と θ に対する成功確率と害リスクの提示を中核に置く。

では、その事後分布はどのように得られるのだろうか。介入効果 τ について利用可能な情報と、今回のデータがもつ情報を統合することで得られる。この関係は、次のように表される。

$$\Pr(\tau | data) \propto \Pr(data | \tau) \times \Pr(\tau)$$

右辺の $\Pr(\tau)$ は事前分布であり、介入効果 τ がどの程度の値を取りそうかという見通しを確率分布として表現したものである。この見通しの根拠としては、過去の研究結果や類似事業の実績、行政データ、専門家予想などが用いられる場合がある。一方で、そのような情報が特にならない場合や、効果の大きさについて特定の予想を置きたくない場合には、幅広い可能性を認める事前分布を用いることもできる⁴。もう一方の $\Pr(data | \tau)$ は尤度であり、「もし τ がこの値だと仮定したら、今回のデータが観測されるのはどれくらい尤もらしいか」を表す。事後分布は、これら二つの情報を統合した結果として得られる。

以上を踏まえ、次節では、前節の仮想数値例（状態 1~4）をそのまま用い、信頼区間による分類が、事後分布上ではどのような確率として現れるかを確認する。

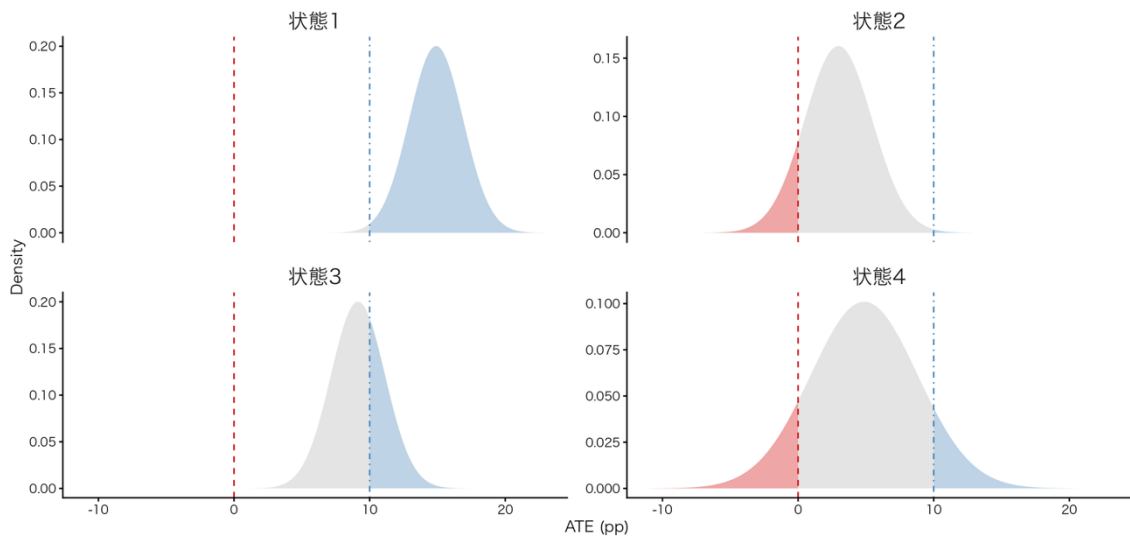
3.2. 事後分布による 4 状態の把握

図 3 は、前節と同じ 4 つの数値例（点推定値 $\hat{\tau}$ と標準誤差 SE ）について τ の事後分布を描いたものである⁵。図 3 では、0 未満の面積が害リスク $\Pr(\tau < 0 | data)$ 、 θ 以上の面積が成功確率 $\Pr(\tau \geq \theta | data)$ に対応し、残り（0 と θ の間）が「十分未満だが非害」に残る確率を表す。したがって図 3 は、図 2 で位置関係として整理した四つの状態を、同じ 0 と θ の境界に対する確率として示したものである。

⁴ 「事前分布」という名称は、過去研究や専門家予想などの具体的な事前情報を必ず反映しなければならないという印象を与えやすい。しかし、実際には、そのような情報が特にならない場合や、効果の大きさについて特定の予想を置きたくない場合であっても、事後分布を得ることはできる。本稿で強調しておきたいのは、こうした場合であっても、十分性や安全性に関する問いを確率として表現できるという点である。

⁵ 事後分布の計算では、推定量の正規近似として $\hat{\tau} | \tau \sim Normal(\tau, SE^2)$ を置き、事前分布として $\tau \sim Normal(0, 25^2)$ を用いた。これは、効果が 0 付近にありつつも ±数十 pp（%ポイント）程度までは取りうるという、幅を持たせた仮定である。

図 3 各状態に対応する事後分布



前節の 4 つの数値例（状態 1～4）について、事後分布から $\Pr(\tau \geq \theta | data)$ と $\Pr(\tau < 0 | data)$ を計算すると表 2 の通りとなる。この表が示しているのは、前節で「未確定」として残った論点が、事後分布上ではどの程度の確率として残っているかである。たとえば状態 3 では、信頼区間の整理は「0 未満は排除できるが、 θ 以上かどうかは未確定」で止まる。しかし事後分布を用いると、 θ 以上である確率は約 33%である一方、0 未満の確率はほぼ 0 であることが分かる。これは、同じ「未確定」でも、 θ を超える見込みがどの程度あるのかを確率として判断材料に加えられることを意味する。

状態 4 ではこの違いがより典型的に現れる。信頼区間は 0 も θ もまたぐため、十分性も安全性も未確定であることまでは明確になる。しかし、その先の「 θ 以上である確率」や「0 未満である確率」を、区間の提示だけから読み取ることはできない。事後分布を用いれば、同じ境界（0 と θ ）に対して、それぞれを 9.7%、10.9% のように確率として示せる。したがって議論は、「両方が未確定」という整理にとどまらず、十分性がどの程度見込めるか、害がどの程度懸念されるかを確率として踏まえた評価へ移りやすくなる。

次章では、この情報がどのように意思決定の議論を組み替える（あるいは確信度を補強する）のかを実際の事例を用いて確認する。

表 2 信頼区間による状態整理と事後分布が与える確率

状態	前節の設定 ($\hat{\tau}$, SE , 95% CI)	頻度論の整理	ベイズ： $Pr(\tau \geq \theta data)$	ベイズ： $Pr(\tau < 0 data)$	参考： $Pr(0 \leq \tau < \theta data)$
1	$\hat{\tau} = 15, SE = 2,$ $CI = [11.1, 18.9]$	十分性：支持 安全性：支持	99.3%	0.0% (ほぼ 0)	0.7%
2	$\hat{\tau} = 3, SE = 2.5$ $CI = [-1.9, 7.9]$	十分性：非支持	0.2%	11.6%	88.1%
3	$\hat{\tau} = 9.2, SE = 2$ $CI = [5.3, 13.1]$	十分性： 未確定 安全性：支持	33.3%	0.0% (ほぼ 0)	66.7%
4	$\hat{\tau} = 5, SE = 4$ $CI = [-2.8, 12.8]$	十分性： 未確定 安全性： 未確定	9.7%	10.9%	79.4%

4. 事例 1：職業訓練プログラムの効果検証と政策判断

本章では、トルコで実施された職業訓練プログラムの RCT を素材として、政策判断に必要な二つの問い——十分性 (θ を満たすか) と安全性 (0 未満を避けられるか) ——を、標準的な信頼区間とベイズの事後分布でそれぞれどう扱えるかを対比して示す。とくに、信頼区間が 0 と θ の両方を含む状態 4 に該当する就労アウトカムを取り上げる。状態 4 では、頻度論の整理は「十分性も安全性も未確定」という形で論点の構造を明確にする一方、その先の判断に必要な情報が不足しやすい。ここに事後分布を用いると、同じ二つの問いは成功確率と害リスクという二つの確率として表現でき、議論の焦点が定まりやすくなる。以下では、この事例を通じて、議論がどこで止まり、どの情報が加わると何が論点として前に出てくるのかが、より具体的に把握できることを示す。

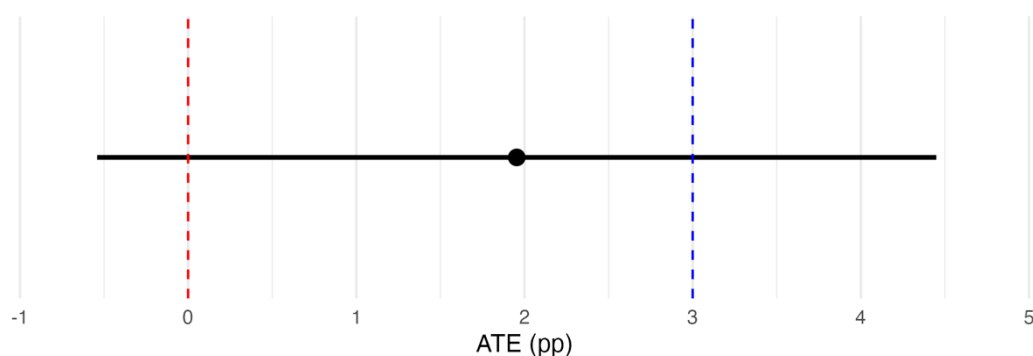
事例として用いる職業訓練プログラムの RCT は、失業者を対象とした職業訓練の効果を検証するために実施されたものである。応募者に対して訓練への参加機会が個人単位でランダムに割り当てられ、参加を許可された介入群と参加できなかった統制群が無作為に形成されている。主なアウトカムは、一定期間後に雇用されているかどうかを示す就労の有無であり、二値アウトカムとして観測される。実データとしては、個人ごとに割付（訓練参加機会の付与）と就労の有無が記録されており、本章では欠損のない 5,497 人を対象に、介入群と統制群の就労割合の差を効果の推定量として用いる。本 RCT の詳細は Hirshleifer et al. (2016) を参照のこと。

本事例において、政策判断者は職業訓練プログラムが有効な取り組みであり、今後も実施を継続する価値があるかを見極めたいと考えているとする。RCT によって、就労割合を 3pp 以上押し上げる見込みがあるという結果が得られるのであれば、継続という判断が支持される。また、意思決定上の説明責任として負の効果が生じる可能性も踏まえて判断する必要がある。

4.1. 推定結果

まず就労アウトカムについて、点推定値と 95%信頼区間を推定し、 $\kappa = 0$ と $\theta = 3$ という二つの境界に重ねて状態を判定する。実データを用いた推定結果は、点推定値が 2.00、95%信頼区間が $[-0.50, 4.40]$ である (図 4)。

図 4 就労アウトカムの推定結果 (95%信頼区間)

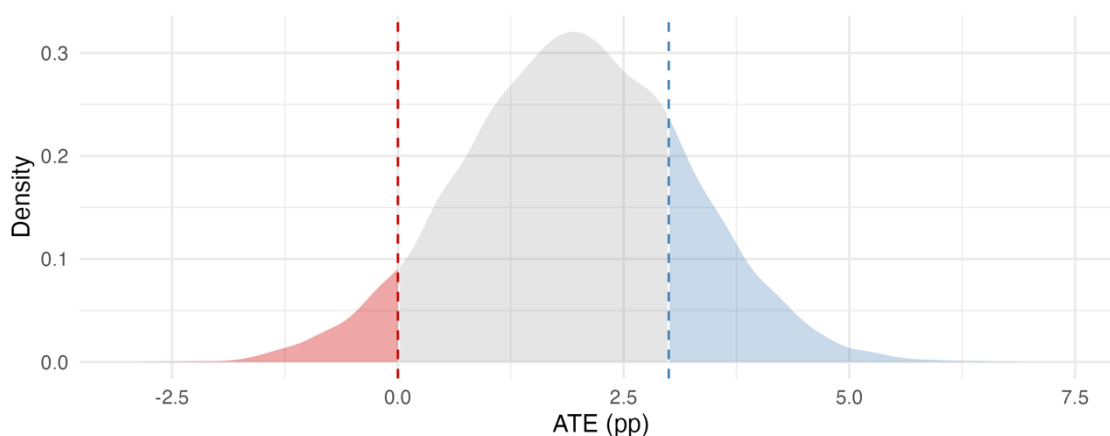


この区間は 0 も $\theta = 3$ も含むため、本ケースは十分性も安全性も未確定な状態 4 に該当する。つまり、(i) 平均として悪化している可能性 (0 未満) が排除できないこと、(ii) 採択に値する改善 (θ 以上) の可能性も排除できないこと、が同時に残る。一方で、信頼区間の提示だけでは、 θ 以上である確率や 0 未満である確率を直接には示せないため、状態 4 のまま判断が止まりやすい。

そこで、同じデータから τ の事後分布をベイズ推定すると図 5 のような結果を得ることができる。なお、ここでは効果の大きさについて幅広い可能性を認める事前分布を用いている。まず、得られた事後分布の要約は、事後平均が 1.92、95%信用区間が $[-0.61, 4.34]$ である。そして、 τ が 0 と θ によって区分される各領域の値をとる確率は以下のように求まる。

- ・ 成功確率 (十分性) : $\Pr(\tau \geq 3 \mid data) = 19.4\%$
- ・ 害リスク (安全性) : $\Pr(\tau < 0 \mid data) = 6.6\%$
- ・ 中間域 : $\Pr(0 \leq \tau < 3 \mid data) = 74.0\%$

図 5 就労アウトカムの事後分布



信頼区間による整理は状態 4（十分性も安全性も未確定）というところで止まるのに対し、事後分布を用いると、 θ 以上である確率と 0 未満である確率を示せる。これにより意思決定のための検討は、「未確定」という整理にとどまらず、成功確率と害リスクを踏まえた判断へと進めやすくなる。

4.2. 判断に向けた検討

信頼区間であっても事後分布であっても、推定結果だけで採否が自動的に決められるわけではなく、最後は意思決定者が何を重視するかという判断に委ねられる点は共通である。両者の違いは、その判断に必要な材料をどのような形で提示できるかにある。信頼区間は「十分性と安全性が未確定である」という状況整理を与え、事後分布は同じ問いを成功確率と害リスクという確率として示す。以下では、この追加情報によって論点がどのように分かれ、どこで価値判断を置く必要が生じるのかという観点から、採否判断に向けた検討がどのように組み立てられるかを整理する。

第一に、成功確率 19.4%をどのように評価するかという問いが前面に現れる。3pp 以上の改善が見込める確率が約 2 割であるという結果を、「低い」と捉えるのか、それとも「不確実性の下では一定の可能性が残っている」と捉えるのかは、介入コスト、代替政策の有無、政策目的の優先度などに依存する。例えば、この職業訓練が高コストで、実施すると他の有望な施策の予算を圧迫するような状況であれば、成功確率が約 2 割という見込みに賭けるのは割に合わない、という判断になりやすい。他方で、比較的低コストで段階導入や中止が容易であれば、2 割でも「試す価値がある」と評価し得る。重要なのは、判断の焦点が「有意か否か」や「未確定か否か」ではなく、成功確率という水準そのものに置き直される点である。

第二に、害リスク 6.6%をどの程度許容できるかという問いが明確になる。平均として悪化する可能性がゼロでない以上、安全性が完全に確保されたとは言えない。しかし、その確率が約 7%であることが示されることで、害の可能性を理由に一律に採択を否定するのではなく、どの条件であれば許容できるのか、あるいはどの追加情報があれば判断を進められるのかといった検討に接続しやすくなる。例えば、まずは小規模に段階導入し、就業アウトカムの推移を定期的に確認したうえで、一定の基準を下回る兆候が見られた場合には速やかに縮小・中止する、という運用は「約 7%」という害リスクの扱い方として具体的に議論に載る。他方で、政策が一度走り出すと止めにくい（可逆性が低い）状況や、悪化が重大な損失につながり得る状況であれば、同じ約 7%でも許容できないと判断し、追加検証や設計変更を優先するという整理も成り立つ。このように、実施規模の調整、段階的導入、追加的なモニタリングや中止条件の設定といった運用上の選択肢が、具体的な論点として浮かび上がる。

このように、事後分布は、従来型の整理によって「未確定」として一括りにされていた状態 4 の内部を、成功確率と害リスクという二つの軸に分解する。これにより、判断は「結論が出ない」という段階で止まるのではなく、「どの確率をどのように評価するか」「どの価値判断を引き受けるか」という次の段階へと進むことが可能になる。ベイズ的推定の付加価値は結論を与えることではなく、判断に必要な論点と責任の所在を、確率という形で可視化する点にある。

5. 事例 2 : UCT における悪影響の検証と政策判断

本章では、政策判断においてしばしば中心となる「悪化の許容可能性」の問いを扱う。これは典型的には、介入の副作用アウトカムをめぐる問いである。ここで前面化する関心は、介入が望ましい改善を十分にもたらすか（充分性）という点ではなく、悪化がどの程度までなら許容できるかという安全性の側にある。そのため本章では、充分性の境界は判断条件としては用いず、安全性の評価に焦点を絞る。さらに安全性についても、悪化が完全に避けられるかという確認にとどまらず、どの程度までの悪化を許容するのかという許容上限を明示して検討する⁶。以下では、ケニアで実施された無条件現金給付（Unconditional Cash Transfer: UCT）の RCT を素材として、こうした場面においても事後分布によって論点がどのように整理されるのかを、事例を通じて具体的に示す。

事例として用いる UCT の RCT は、ケニア農村部の困窮世帯を対象として実施されたものであり、世帯の資産や消費といった厚生水準に好ましい影響を与えることを目的としていた。村内で UCT の受給対象となる世帯はランダム割付によって決められており、介入群 471 世帯と統制群 469 世帯が形成されている。本章では、この 940 世帯から得られたベースラインおよびエンドラインのデータを用いる。本 RCT の詳細は Haushofer and Shapiro (2016) を参照のこと。

本事例において、政策判断者は、UCT が家計の厚生水準に好ましい影響を与える政策である可能性を認めつつも、負の影響が生じていないかを確認したうえで採否を判断したいと考えているとする。現金給付は資産形成や消費平準化に資する有効な政策であると期待される一方で、奢侈品・嗜好品の消費増加や生活行動の望ましくない変化を招くのではないかと、という批判も多い。そこで本章では、主アウトカムに関する検証を進めた後に、副作用が生じていないかを検討している状況における判断を扱う。具体的には嗜好品消費としてアル

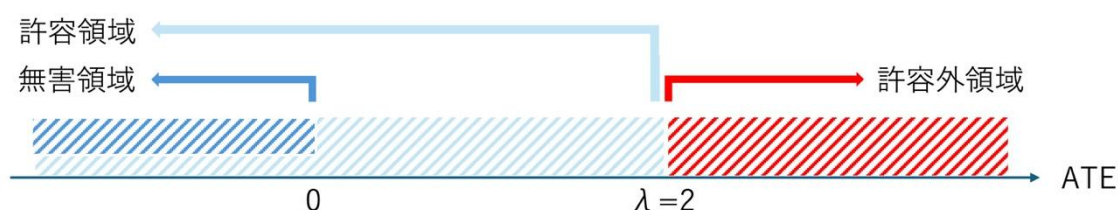
⁶ 本稿の枠組みでは、安全性は一般に境界 κ を置いて評価でき、 κ は 0 に限らず任意の値を取りうる。本章の「許容上限」は、この κ を悪影響の上限として具体化したものである。なお 0 は、許容上限とは別に、「悪化が生じていない（あるいは改善している）」ことを示す参照点として併置する。安全性をこのように上限と参照点に分けて読み分けること自体は、充分性・安全性の一般の議論の中でも理屈の上では可能であるが、本稿では議論を簡潔にするためこれまで明示的には扱ってこなかった。

コール消費額を取り上げる。

5.1. 悪影響の許容可能性をめぐる判断枠組み

本事例では、副作用アウトカムとしてアルコール消費額を扱う。本節では、安全性の評価に焦点を絞り、その安全性を許容上限 λ と 0 によってより細かく読み分ける。そこで、アルコール消費額の変化を次の三つの領域に区別して考える（図 6）。

図 6 悪影響の許容可能性に関する領域区分



第一に、無害領域である。これは、アルコール消費額が増加していない、あるいは減少している領域を指す。この領域に推定結果が位置づけられる場合、少なくとも副作用が生じているとは言えず、政策判断上の懸念は最も小さい。

第二に、許容領域である。これは、アルコール消費額が増加している可能性を含みつつも、その増加が実務的に許容可能な範囲にとどまっている領域を指す。本事例では、政策担当者が「この程度までであれば、他の便益とのトレードオフとして受け入れうる」と考える水準として $\lambda=2(\text{USD})$ を設定する。この領域に推定結果が位置づけられる場合、悪影響が生じうることを踏まえつつも、条件付きでの採択や運用上の工夫が検討対象となる。

第三に、許容外領域である。これは、アルコール消費額の増加が λ を超える領域であり、副作用として看過できない水準に達していると判断される範囲である。この領域に相当する可能性が大きい場合、政策の採択や継続は難しく、設計の見直しや中止が現実的な選択肢として浮上する。

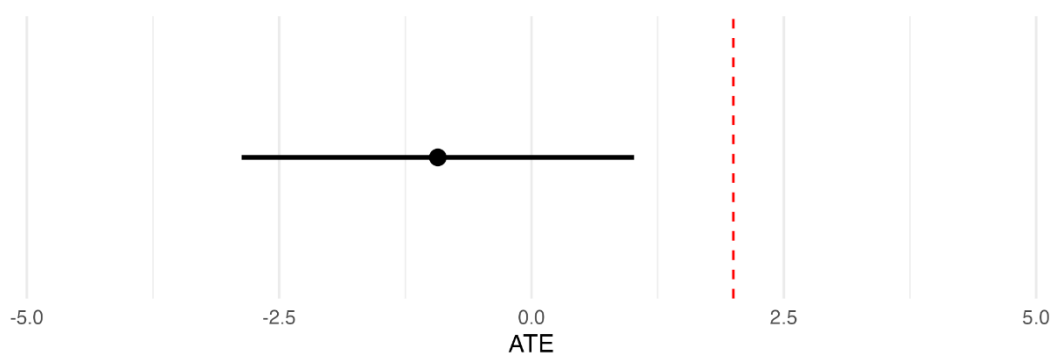
以下では、この整理を前提として、アルコール消費額に関する推定結果を提示し、信頼区間ではどこが未確定として残るのか、事後分布によってどの情報が加わると論点がどう整理されるのかを確認する。

5.2. 推定結果

まず、アルコール消費額について、介入群と統制群の平均差を平均介入効果 τ の推定値と

して得る。頻度論の標準出力は点推定値と 95%信頼区間である。本事例の推定結果は、点推定値が $\tau = -0.93(\text{USD})$ 、標準誤差が $SE = 0.99$ であり、95%信頼区間は $[-2.86, 1.02]$ である (図 7)。

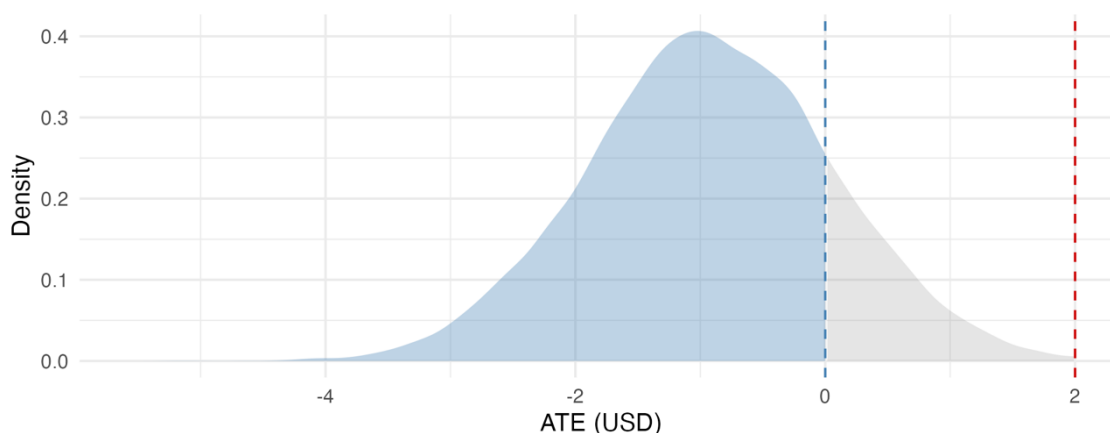
図 7 アルコール消費額の推定結果 (95%信頼区間)



この信頼区間を図 6 の境界 (0 と $\lambda = 2$) に重ねて読むと、区間は 0 をまたいでいる。したがって、アルコール消費が「増加していない」可能性は残る一方で、「増加している」可能性も排除できない。しかし同時に、信頼区間の上限は 1.02 であり、許容外領域の境界である $\lambda = 2$ を超えていない。少なくとも「2USD を超える増加」という含意は、95%信頼区間の提示だけからは支持されにくい。

次に、同じデータから τ の事後分布を得て、同じ境界 (0 と $\lambda = 2$) に対する確率として読む。ここでも、事例 1 と同様に効果の大きさについて幅広い可能性を認める事前分布を用いている。事後分布の要約は、事後平均が -0.92 (USD)、95%信用区間が $[-2.86, 1.02]$ である。区間の幅は頻度論の 95%信頼区間とほぼ同じであり、ベイズの付加価値は「区間が狭くなる」ことにあるのではない。

図 8 アルコール消費額の事後分布



一方で事後分布からは、許容上限を超える悪化の確率と、無害である確率を直接に得られる。許容外領域に入る確率は $\Pr(\tau > \lambda | data) = 0.00142$ である。また無害領域に入る確率は $\Pr(\tau \leq 0 | data) = 0.826$ である。したがって、許容上限の範囲内にとどまる確率は $\Pr(\tau \leq \lambda | data) = 1 - 0.00142 = 0.99858$ であり、そのうち「増加してはいるが許容上限は超えない」部分の確率は $\Pr(0 < \tau \leq \lambda | data) = 0.99858 - 0.826 = 0.1726$ となる。

5.3. 推定結果と判断に向けた検討

前節の推定結果を許容可能性の枠組みで読むと、まず確認すべき問いは「許容外領域 ($\tau > \lambda$) に入っている可能性がどの程度あるか」である。本事例では、頻度論の95%信頼区間の上限は 1.02 であり、許容上限 $\lambda = 2$ を超えていない。したがって、95%信頼区間の提示だけから見ても、「許容外の水準 (2USD 超の増加) が生じている」という含意は支持されにくい。少なくとも、許容外と判断して採択や継続を否定する根拠は弱い。

ただし、ここまでの整理は「許容外かどうか」に関するものであり、許容範囲内で悪化が起きうるかどうかは別の論点として残る。頻度論では区間が 0 をまたぐため、「増加がない」と言い切ることはできないが、増加が起きうる確率を数として示すことはできない。

これに対してベイズの事後分布からは、許容外となる確率 $\Pr(\tau > \lambda | data)$ が極めて小さいことに加え、参考として「許容範囲内ではあるが悪化する」確率 $\Pr(0 < \tau \leq \lambda | data)$ も得られる。したがって本事例の判断は、「許容外の可能性はほぼ排除できる」という見通しを軸にしつつ、許容範囲内に残る悪化リスクをどの程度引き受けるか、という形で具体化される。

本章では、充分性と安全性の二つの問いを整理してきた見方のうち、副作用アウトカムに対

応して安全性の側だけを取り出し、 0 と λ を用いてその安全性をより細かく読み分けた。これにより、許容外の可能性と許容範囲内に残る悪化リスクを確率として並べて提示でき、判断の根拠を説明可能な形に落とし込めることを示した。

6. まとめ

本稿では、インパクト評価の結果が政策判断においてどのように用いられうるのかを、十分性・安全性・許容可能性という判断上の関心に照らして検討してきた。標準的な提示形式である点推定値と信頼区間は、境界を重ねて読むことで、判断に必要な論点を一定程度まで整理することができる。しかし同時に、境界をまたぐ状況では「何が未確定なのか」は明らかになっても、その不確実性をどう引き受けるかという次の問いには答えきれないことも確認された。

事後分布を用いた提示は、こうした局面において、判断を代替する結論を与えるものではない。その役割は、境界や許容範囲に対して推定結果を重ね、どの可能性がどの程度残っているのかを確率として可視化する点にある。これにより、判断は「採択できるか否か」で止まるのではなく、「どのリスクをどの程度引き受けた判断なのか」を説明できる形へと進む。本稿で示した枠組みは、特定の事例に限られたものではなく、判断の関心に応じて境界や範囲を設定し、その上に推定結果を重ねて読むという、汎用的な発想として位置づけられる。

参考文献

Haushofer, J., & Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. *The Quarterly Journal of Economics*, 131(4), 1973-2042.

Hirshleifer, S., McKenzie, D., Almeida, R., & Ridao-Cano, C. (2016). The impact of vocational training for the unemployed: experimental evidence from Turkey. *The Economic Journal*, 126(597), 2115-2146.