

## 숫자양식과 빈도범위가 베이스 추론 과제에서 위험판단과 추론에 미치는 영향\*

이 현 주                      이 영 애†  
이화여자대학교 심리학과

위험정보를 확률이나 빈도양식으로 제시하고 질병으로 사망할 확률(기저율)에 대한 위험을 판단하게 하고 양성판정을 받은 사람이 질병에 걸렸을 확률(사후확률)에 대한 위험판단과 추론의 정확성을 비교하였다. 베이스 추론 과제를 사용한 연구 1에서 숫자양식의 효과가 관찰되었다. 참가자들은 위험이 확률보다는 빈도로 제시될 때 더 위험하다고 판단하였고 질병에 걸렸을 확률을 더 정확하게 추론하였다. 빈도의 범위가 좁을 때보다 넓은 때 더 위험하다고 판단하는 효과는 관찰되지 않았다. 분석적 사고체계가 위험판단에 미치는 영향을 검토하려고 사후확률을 계산하는 조건과 계산하지 않는 조건을 비교하였다. 숫자양식의 효과는 여전히 관찰되었다. 연구 2는 기저율과 사후확률의 크기에 따라 숫자양식 효과와 빈도범위 효과가 달라지는지 알아보았다. 숫자양식의 효과는 기저율과 사후확률의 크기에 상관없이 모든 조건에서 관찰되었다. 위험한 사건이 발생할 확률의 높고 낮음에 상관없이 빈도로 제시되었을 때 참가자들이 더 위험하다고 판단하였다. 그러나 빈도범위 효과는 기저율이 낮은 조건에서만 발견되었다. 본 연구의 결과들을 이중처리체계 이론과 관련시켜 논의하였다.

주제어 : 숫자양식, 빈도범위, 확률, 빈도, 베이스 추론, 위험판단

\* 이 연구는 한국학술진흥재단 기초연구과제지원사업(KRF-2007-321-H00022)의 지원에 의해 수행되었다.

이 논문을 읽고 세심한 지적을 해준 심사위원들에게 감사드린다.

† 교신저자: 이영애, 이화여자대학교 심리학과 연구세부분야: 판단과 결정,  
E-mail: yalee@ewha.ac.kr

질병으로 사망할 가능성처럼 위험한 일이 발생할 가능성은 빈도(frequency)나 확률(probability) 등의 숫자로 제시된다. 위험의 판단은 여러 요인들의 영향을 받을 수 있으나, 위험관련 정보가 어떤 숫자양식으로 제시되는지에 따라 달라진다. Slovic, Monahan과 MacGregor (2000)는 임상심리학자들에게 정신질환자의 폭력행동 가능성을 판단하게 하였다. 임상심리학자들은 정신질환자의 위험성이 빈도로 제시되는가 아니면 확률로 제시되는가에 따라 위험정도를 다르게 판단하였다. 퇴원 후 정신질환자가 폭력행동을 할 가능성이 ‘100 명 중 20 명(빈도양식)’이라고 했을 때에는 폭력행동을 할 ‘확률이 20%(확률양식)’라고 했을 때보다 위험이 더 크다고 지각하였다. 위험정보를 빈도로 제시받은 임상심리학자들은 정신질환자의 상태를 더 위험하게 생각하여 환자의 퇴원을 반대하는 결정을 더 많이 내렸다. 임상심리학자와 같은 전문가들도 숫자양식에 따라 위험지각 및 판단이 영향을 받는다.

빈도로 위험정보를 제시할 때는 범위를 정해야 한다. 같은 위험도 어떤 범위에서 제시되느냐에 따라 위험이 다르게 느껴질 수 있다. Yamagish (1997)는 암이나 자동차 사고처럼 잘 알려진 원인별로 사망 가능성을 여러 범위에서 제시하고 위험을 평가하도록 하였다. 예를 들어, 참가자들은 ‘10,000 명 중에서 2,414 명이 암으로 사망한다.’는 것이 얼마나 위험하게 지각되는지를 평가하였다. 사람들은 사망 가능성이 같아도 범위에 따라 위험을 다르게 지각하였고 사망 가능성이 다른 때조차 큰 범위로 제시된 작은 가능성(예, 10,000 명 중 1,286 명 사망)을 작은 범위로 제시된 큰 가능성(예, 100 명 중 24.14 명 사망)보다 더 위험하다고 판단했다.

본 연구는 숫자양식 효과와 빈도범위 효과를 이중처리체계(dual process theory) 관점으로 접근한다. 연구자들은 인간의 사고체계가 두 체계로 되어 있다고 주장한다(예: Chaiken & Trope, 1999; Epstein, 1994; Kahneman & Frederick, 2002; Slovic, 1996). 정보를 느리게 처리하며 논리규칙을 따르며 분석적 사고를 담당하는 분석체계(analytic system)와 정보를 빠르게 처리하며 감정을 동반한 심상(image)으로 현실을 약호화하는 경험체계(experiential system)이다. 두 체계는 독립적 또는 상호작용적으로 판단과 결정에 영향을 준다.

Epstein은 흰 구슬과 빨간 구슬이 섞여 있는 두 쟁반 중에서 빨간 구슬을 뽑으려면 이기는 과제를 제시하였다(Denes-Raj & Epstein, 1994; Denes-Raj, Epstein, & Cole, 1995; Kirkpatrick & Epstein, 1992; Pacini & Epstein, 1999). 사람들은 두 쟁반에서 빨

간 구슬의 비율이 같더라도 절대수가 적은 쟁반(예, 구슬 10개 중 빨간 구슬 1개)보다 절대수가 더 많은 쟁반(예, 구슬 100개 중 빨간 구슬 10개)을 선호하였다. 심지어 빨간 구슬의 비율이 크지만 절대수가 적은 쟁반(예, 구슬 10개 중 빨간 구슬 1개)과 빨간 구슬의 비율이 작더라도 절대수가 더 많은 쟁반(예, 구슬 100개 중 빨간 구슬 7개) 중에서도 절대수가 많은 쟁반을 선호하였다(Denes-Raj & Epstein, 1994). 이 결과에서 볼 수 있는 것처럼 작은 확률로 발생하는 사건 가능성에 대하여 작은 수(예, 10의 1)보다 큰 수의 형태(예, 100의 10)로 제시했을 때 가능성을 더 크게 지각하는 것을 **비율 편향(ratio-bias)** 현상이라고 한다. Finucane, Peters와 Slovic(2003)은 Epstein(1994)의 참가자들이 어떤 쟁반을 선택할지 결정할 때 ‘분자(빨간 구슬의 수)를 주목하고, 분모(쟁반의 전체 구슬 수)는 무시하는 심적 방략’을 사용하였을 가능성을 제기하였다. 사람들은 빨간 구슬의 비율이 어느 쪽이 더 많은지 머리 속으로 알고 있음에도 불구하고 막상 선택을 해야하는 상황에서는 마음속으로 빨간 구슬을 더 많이 떠올릴 수 있는 쪽을 선호하였다 (경험체계의 작용).

Slovic 등(2000)은 숫자양식 효과를 보인 참가자들이 사후질문에서 조건에 따라 다른 심상을 떠올렸음을 발견하였다. 정신질환자의 폭력행동의 가능성이 확률로 제시된 조건의 참가자들은 다른 사람에게 해를 끼치지 않을 사람의 심상이 떠올랐다고 했던 반면, 빈도로 제시된 조건의 참가자들은 무서운 다수의 정신질환자들의 심상이 떠올랐다고 응답하였다(Finucane 등, 2003). 위험에 대한 판단은 합리적으로 이루어져야 하지만 선행연구 결과들(Slovic & Monahan, 1995; Slovic 등, 2000; Yamagishi, 1997)을 검토해볼 때, 정보 제시양식 등의 영향을 받는다. 숫자양식에 따른 위험판단의 차이는 분석체계와 경험체계의 결과로 보인다. 본 연구에서는 건강위험에 대한 숫자정보를 포함하는 의학적 베이스 추론 과제를 택하여 숫자양식의 효과와 빈도범위효과를 이중처리체계 관점에서 검토할 것이다.

판단 및 의사결정 분야의 연구자들(예, Eddy, 1982; Kahneman & Tversky, 1972; Gigerenzer & Hoffrage, 1995)은 **베이스 추론 과제(Bayesian inference task)**를 개발하였다. 이 과제는 참가자들에게 한 관찰을 근거로 상호 배타적이고 포괄적인 두 가설들(mutually exclusive and exhaustive hypothesis)에 대한 수치를 확률 또는 빈도로 추론하도록 한다. 베이스 과제는 확률 법칙에 따라 추론하는 인간의 능력을 검토하는데 중요한 인지 과제로 알려져 있다. 예를 들어, 질병의 사전확률(prior probability)을

아는 상태에서 새로운 자료(예, 증상)를 고려하여 그 질병이 발생할 확률, 즉 사후 확률(posterior probability)을 추정한다. 베이스 추론 과제 의 예로 유방암 진단문제를 보자.

정기 건강검진을 받은 40세 여성의 경우, 유방암에 걸릴 확률은 1% 이다. 유방암에 걸린 여성이 양성판정을 받을 확률은 80 %이다. 유방암에 걸리지 않은 여성이 양성판정을 받을 확률은 9.6 %이다. 이 연령의 한 여성이 정기 건강검진에서 양성판정을 받았다. 그 여성이 실제로 유방암에 걸렸을 확률은 얼마인가?

양성판정을 받은 여성이 실제로 유방암인 사후확률을 구하는 문제에서 기저율(base rate 또는 사전확률), 참-양성판정율(hit rate) 그리고 거짓-양성판정율(false-alarm rate)을 모두 고려해야 한다. 그러나 숫자정보가 확률로 제시되면 전문가든 일반인이든 베이스 추론을 잘하지 못한다(예: Gigerenzer, 1996; Hoffrage & Gigerenzer, 1998). 100명의 의사들에게 유방암 진단문제를 제시했을 때 사후확률을 정확하게 추론한 의사는 5%에 불과하였다(Eddy, 1982). Kahneman과 Tversky(1972, p. 450)는 베이스 추론 과제에서 사람들이 기저율을 무시하고 있음을 발견하고 ‘인간은 베이스 추론을 잘 하지 못한다(he is not Bayesian at all)’고 결론을 내렸다. 사람들이 베이스 추론 과제를 잘 해결하지 못하는 이유는 무엇인가? 숫자를 표상하는 방식 때문이다(Chase, Hertwig & Gigerenzer, 1998; Gigerenzer와 Hoffrage, 1995). 사람들이 베이스 추론을 할 능력이 없는 것이 아니라 숫자정보를 적절하게 표상을 하지 못해 해결하지 못할 뿐이다. 예를 들어, 아라비아 숫자 959와 로마숫자 CMLIX는 같은 값을 나타내지만, 숫자양식에 값들을 계산하는 절차가 다르다. 유방암 진단문제에서 빈도로 제시되면 계산이 쉬워 사후확률을 잘 추론한다. 베이스 추론 과제에서 숫자 정보를 빈도로 제시했을 때 6학년 아동의 53%가 풀 수 있었다 (Zu & Gigerenzer, 2006).

선행연구들(Gigerenzer & Hoffrage, 1995; Kahneman & Tversky, 1972; Slovic & Monahan, 1995; Slovic, Monahan & MacGregor, 2000; Yamagish, 1997)은 위험정보에 대한 추론과 위험판단을 별개의 문제로 다루었다. 질병에 대한 의학적 결정을 할 때

사람들은 숫자정보로 추론하면서 위험을 판단한다. 추론과 판단은 별개의 문제가 아니다. 본 연구에서는 사람들이 위험에 관한 숫자정보로 어떻게 추론하는지 그리고 그 정보로 위험을 어떻게 판단하는지를 함께 다룬다. 개관된 Gigerenzer와 동료들의 연구는 숫자양식을 바꾸면 정확한 추론이 가능해지고 그에 따라 베이스 추론 과제에서 위험판단과 결정이 향상될 것을 시사한다.

본 연구의 첫째 목적은 베이스 추론과제에서 위험에 관한 숫자정보를 확률 또는 빈도로 제시했을 때 위험판단이 달라지는지 비교하는 것이다(연구 1). 본 연구는 기저율과 사후확률의 두 위험정보를 다룬다. 한 위험정보는 기저율인데 질병에 걸려 사망할 가능성이 있다. Slovic 등(2000)의 결과가 타당하다면 베이스 추론과제에서도 **숫자양식의 효과**가 관찰되어야 한다. Yamagishi(1997)의 결과가 타당하다면 베이스 추론과제에서 **빈도범위의 효과**가 관찰되어야 한다. 본 연구는 또 다른 위험정보인 사후확률을 고려하는 위험판단을 검토할 것이다. Yamagishi(1997)는 기저율 정보만을 다루었다. 본 연구는 베이스 추론과제의 특성상 사후확률이 중요하므로 이 확률을 고려한 위험판단이 숫자양식과 빈도범위에 따라 달라지는지 검증할 것이다. 그리고 이들 결과와 이중처리체계의 관련성을 탐색한다. 본 연구의 두 번째 목적은 확률의 크기에 따른 숫자양식 효과와 빈도범위 효과를 검증하는 것이다. 비율 편향 현상을 고려할 때 위험 가능성이 높을 때와 낮을 때 위험을 판단하는 경향이 다를 것이다. 연구 2에서는 연구 1과 동일한 과제를 사용하되 기저율의 크기와 사후확률의 크기를 고확률과 저확률로 나누어 위험판단을 비교한다. 세 번째 목적은 베이스 추론의 정확성과 관련하여 Gigerenzer와 Hoffrage(1995)가 발견한 결과를 한국 대학생들을 대상으로 반복하는데 있다(연구 1). 위험정보가 확률보다는 빈도로 제시되었을 때 베이스 추론이 더 정확한지를 확인한다.

### 연구 1: 위험판단과 베이스추론은 숫자양식에 따라 다른가?

연구 1은 위험판단과 베이스 추론의 정확성이 확률, 빈도 그리고 빈도 범위에 따라 달라지는지를 검토할 것이다. 분석체계가 활성화되면 사람들은 제시된 숫자 정보들을 바탕으로 합리적인 판단을 내릴 것이다. 가능성의 크기가 같다면 빈도로

제시되든 확률로 제시되든, 아니면 어떤 빈도범위에서 제시되든 제시양식과 상관 없이 위험을 비슷하게 판단할 것이다. 경험체계는 감정을 동반한 심상에 기초하여 정보를 직관적으로 빠르게 처리한다. 이 체계가 활성화되면 숫자들에 대한 심상에 따라 위험판단이 달라질 것이다.

분석체계의 활성화는 또 다른 방식으로 검토된다. Hsee와 Rottenstreich(2004)는 이중처리체계 모형에 근거하여 대상의 가치를 평가하는 심리과정을 계산에 의한 평가와 감정에 의한 평가로 구분하였다. 두 종류의 점화과제(priming task)를 구성하고 계산 또는 감정을 선택적으로 점화시키자 대상에 대한 평가가 달라졌다. 본 연구 1은 참가자가 사후확률을 직접 계산하여 분석적 사고체계의 활성화가 위험판단에 영향을 주는지 검증한다. 분석적 사고체계가 활성화되면 합리적인 판단을 하게 된다. 따라서 위험가능성이 같다면 제시양식이 다르더라도 위험판단은 비슷할 것이다. 즉 숫자양식의 효과 및 빈도범위 효과가 관찰되지 않아야 한다.

## 방 법

### 참가자

서울 소재 대학의 학생 437명이 본 연구 1에 참가하였고 소정의 보상을 받았다.

### 설계

숫자양식에 따른 위험판단과 추론의 정확성을 측정하려고 베이스 추론 과제를 확률양식과 두 빈도양식으로 제시하였다. 확률조건은 숫자정보를 백분율로 제시하였으며, 빈도조건은 Yamagiah(1997)의 연구와 비교를 위해 전체 100명인 좁은 범위와 10,000명인 넓은 범위로 제시하였다. 숫자정보는 확률과 좁은 빈도범위(FNa) 그리고 넓은 빈도범위(FWi)로 제시되었다(표 1). 숫자양식은 참가자 내 변인이었고 계산 여부는 참가자 간 변인이었다. 사후확률에 대한 위험판단은 참가자가 사후확률을 계산하는 조건과 계산할 필요 없는 비계산 조건으로 나누어졌다. 사후확률에 대한 위험판단은 숫자양식(3)과 계산여부(2)의 혼합설계였다.

표 1. 연구 1과 2에서 조건별 숫자정보의 종류와 그 값

제시양식	정보의 종류	연구 1 및 연구 2의 고-확률 조건	연구 2의 저-확률 조건
확률	기저율(사전확률)	12%	2%
	참-양성판정율	모두	모두
	거짓-양성판정율	10%	30%
	정답(사후확률)	57.7%	6.4%
좁은 빈도범위 (FNa)	기저율(사전확률)	100명 중 12명	100명 중 2명
	참-양성판정율	모두	모두
	거짓-양성판정율	88명 중 8.8명	98명 중 29.4명
	정답(사후확률)	20.8명 중 12명	31.4명 중 2명
넓은 빈도범위 (FWi)	기저율(사전확률)	10,000명 중 1,200명	10,000명 중 200명
	참-양성판정율	모두	모두
	거짓-양성판정율	8,800명 중 880명	9,800명 중 2,940명
	정답(사후확률)	2,080명 중 1,200명	3,140명 중 200명

#### 자극 및 절차

연구 1의 자극은 선행연구(Gigerenzer & Hoffrage, 1995)의 자극들과 기본적으로 같으나 몇 가지 점에서 달랐다. 첫째, 연구 1에서는 ‘유방암’과 같은 병명을 사용하지 않고, ‘질병 A’로 명명하여 특정 질병에 대한 참가자의 선입견을 배제하고 사망 가능성의 위험성만을 측정하였다. 둘째, 확률과 빈도양식의 차이만 비교하지 않고 빈도범위 효과를 검증하기 위해 빈도정보를 좁은 범위(100명)와 넓은 범위(10,000명)에서 제시하였다. 셋째, 숫자정보를 참-양성판정율을 제외한 두 종류(기저율, 거짓-양성판정율)로 제한하였다. 베이스 추론 과제의 오류는 거짓-양성판정율을 고려하지 못하여 발생한다(Krynski & Tenenbaum, 2007). 따라서 질병 A에 걸린 사람들이 모두 양성판정을 받을 것이라고 바꾸고(즉, 참-양성판정율 100%), 거짓-양성판정율을 고려하는 여부에 따라 정답이 달라지도록 하였다. 문제를 수정하여 참

가자들이 사후확률을 구하는 공식에 숫자를 대입하기보다 추론하도록 유도하는 동시에 계산을 단순화시켜 계산상 오류를 줄였다. 참가자들에게 연구 1을 결정에 관한 설문조사로 소개하였다. 지시문에 이어 제시된 진술문(예, 좁은 빈도범위 조건)은 다음과 같다.

매년 우리나라 성인남녀의 100명 중 12명이 질병 A에 걸릴 것이다. 이 검사를 받게 되면, 질병 A에 걸린 사람들 모두 양성(+)이라고 판정될 것이다. 그러나 질병 A에 걸리지 않은 나머지 88명 중 8.8명도 양성(+)이라는 판정을 받게 될 가능성이 있다.

검사를 받고 양성(+)이라는 판정을 받은 사람들 중에서 실제로 질병 A에 걸려 있는 사람은 몇 명일까?

(계산조건)

검사를 받고 양성(+)이라는 판정을 받은 사람들 중에서 실제로 질병 A에 걸려 있는 사람은 20.8명 중 12명 이다.

(비계산 조건)

참가자들은 **기저율에 대한 위험**을 판단하였다. 확률조건에서는 ‘매년 우리나라 성인남녀의 12%가 질병 A로 사망한다’고 제시하고 질병 A가 얼마나 위험하다고 느껴지는지를 9점 척도에서 평정하게 하였다(‘1점 전혀 위험하지 않다’와 ‘9점 매우 위험하다’). FNa 조건에서는 ‘매년 우리나라 성인남녀 100명 중 12명이 질병 A로 사망한다’고 하고, Fw 조건에서는 ‘매년 우리나라 성인남녀 10,000명 중 1,200명이 질병 A로 사망한다’고 제시하였다.

기저율에 대한 위험을 평정한 후 참가자들은 숫자정보들을 사용하여 사후확률(양성판정을 받은 중에서 실제로 질병 A에 걸렸을 확률)을 추정하였다(계산조건). 그리고 자신이 추정한 **사후확률을 고려했을 때, 질병 A에 관한 검사에서 양성판정을 받는 것이 얼마나 위험한 상황이라고 느껴지는지**를 판단하였다. 비계산 조건의 참가자는 연구자가 제시한 사후확률 값에 대해 위험을 판단하였다. 본 연구



1의 종속측정치는 기저율에 대한 위험평정 및 사후확률 고려시 양성판정에 대한 위험평정(9점 척도), 그리고 사후확률 추정에 대한 정답률(%)이었다.

모든 참가자(437명)들이 기저율에 대한 위험판단을 하였다. 사후확률 추정에 대한 정답률은 계산조건 참가자(327명)를 대상으로 구하였다. 사후확률을 고려한 양성판정에 대한 위험판단은 비계산조건 전체와, 계산조건 참가자 중 사후확률을 정확하게 추정한 정답자에게서 얻어야 한다. 예비연구(76명)에서 베이스 추론문제에 대한 확률조건의 정답률은 빈도조건의 정답률에 비해 매우 낮았다. 연구 1의 계산조건에서는 정답자를 충분히 확보하기 위해 참가자들을 확률조건에 더 많이 배정하였다(계산조건 327명: 확률 163명, FNa 91명, FWi 73명). 양성판정에 대한 위험판단은 계산조건 참가자 중에서 정답자(114명: 확률 31명, FNa 40명, FWi 43명)와 비계산조건 참가자(110명: 확률 36명, FNa 34명, FWi 40명)를 대상으로 얻었다.

## 결과 및 논의

### 기저율에 대한 위험판단

베이스 추론문제를 푸는 것과 상관없이 모든 참가자들이 기저율에 대한 위험평정을 하였다. 연구 1의 관심은 숫자양식의 효과(즉 확률과 빈도조건 간의 차이) 그리고 빈도범위 효과(즉 빈도범위 조건 간의 차이)를 검증하는데 있었다. 두 빈도범위 조건을 합하여 빈도조건을 만들어 확률조건과 비교하고, 빈도조건 내에서 좁은 빈도범위와 넓은 빈도범위를 비교하는 두 개의 **계획된 사전비교(planned comparison)**를 실시하였다. 숫자양식의 효과를 검증하기 위한 사전비교에서 기저율에 대해 빈도조건의 참가자( $M = 6.64$ ,  $SD = 1.70$ )들이 확률조건의 참가자들( $M = 5.71$ ,  $SD = 1.81$ )보다 위험을 통계적으로 유의하게 높게 평정하였다( $t = 5.241$ ,  $df = 434$ ,  $p = .000$ ). 두 빈도범위 조건(FNa 조건과 FWi 조건)을 비교하였으나 위험평정에서 차이가 나타나지 않았다( $t = 1.250$ ,  $df = 434$ ,  $p = .212$ ). 본 연구 1의 결과는 그림 1에 제시되었다. 연구 2는 빈도범위의 효과를 다시 다룰 것이다.

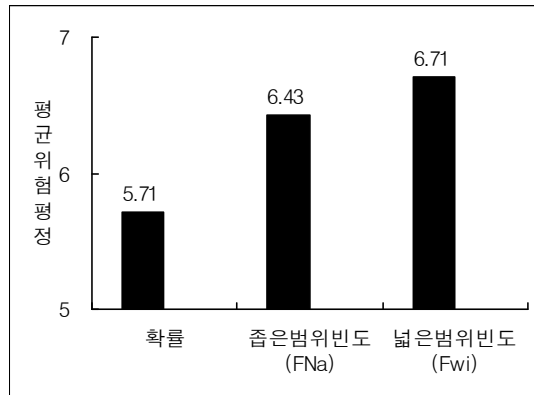


그림 1. 숫자양식별 기저율에 대한 평균 위험평정

### 베이스 추론 정당률

계산조건의 참가자(327명)들이 조건별로 보인 베이스 추론의 정당률을 비교하였다. 빈도범위 조건의 정당률(FNa 조건 44.0 %; FWi 조건 58.9 %)은 확률조건의 수행(19.0 %)보다 높았다[확률조건 대 FNa 조건,  $\chi^2(1, N = 254) = 18.03, p = .000$ ; 확률조건 대 FWi 조건,  $\chi^2(1, N = 236) = 37.27, p = .000$ ]. FWi 조건은 FNa조건보다  $\chi^2(1, N = 164) = 3.62, p = .057$  정당률이 높았으나 통계적으로 유의하지 않았다. 오답을 분석한 결과, 이 결과는 FWi 조건에서 계산상 오류가 적었기 때문인 것으로 해석되었다. 선행연구(Gigerenzer & Hoffrage, 1995)에서처럼 본 연구 1에 참가한 한국 대학생들도 확률보다 빈도로 제시하였을 때 베이스 추론과제에서 사후 확률을 더 정확히 추정하였다.

### 사후확률을 고려했을 때 양성판정에 대한 위험판단

계산조건의 참가자들에게 사후확률 값을 구하게 하고 양성판정에 대한 위험을 평정시켰다. 자료분석에서는 사후확률을 정확하게 추정한 응답자들의 위험평정을

표 2. 연구 1에서 계산조건( $N = 327$ )에서 숫자양식별 정답자 수

제시양식	정답자 수	오답자 수	전체
확률	31	131	163
좁은 빈도범위(FNa)	40	51	91
넓은 빈도범위(FWi)	43	30	73

비교하였다(조건별 정답자수: 확률조건 31명, FNa 40명, FWi 43명). 사후확률을 고려했을 때 양성판정에 대한 위험평정값은 조건별로 다음과 같았다: 확률조건( $M = 6.22$ ,  $SD = 1.60$ ), 좁은 빈도범위 조건( $M = 6.74$ ,  $SD = 1.36$ ), 넓은 빈도범위 조건( $M = 1.76$ ,  $SD = 1.64$ ). 앞서 언급한 것처럼, 사후확률에 대한 위험판단에서는 계산이 포함되므로 숫자양식과 계산여부를 참가자간 변인으로 하는 이원변량분석을 실시하였다. 위험평정에 대한 숫자양식과 계산여부의 상호작용효과가 통계적으로 유의하지 않았다( $F(2, 218) = 0.80$ ,  $MSe = 1.91$ ,  $p = .451$ ). 즉 숫자양식에 따른 위험판단이 계산여부에 의해 달라지지 않았다. 계산여부의 주효과도 나타나지 않았으므로( $F(1, 218) = 0.53$ ,  $MSe = 1.26$ ,  $p = .468$ ) 계산조건과 비계산 조건을 통합하였다. 기저율에 대한 분석처럼 숫자양식의 효과와 빈도범위 효과를 검증하기 위한 두 개의 계획된 사전비교(planned comparison)를 하였다. 확률조건( $M = 6.22$ ,  $SD = 1.60$ )의 참가자들보다 빈도조건( $M = 6.75$ ,  $SD = 1.50$ )의 참가자들이 더 위험하다고 판단하였다( $t = 2.343$ ,  $df = 221$ ,  $p = .020$ ). 두 빈도범위 조건에서 위험평정의 차이를 비교하였으나 통계적으로 유의하지 않았다( $t = .064$ ,  $df = 221$ ,  $p = .949$ ). 사후확률에서도 빈도범위 효과는 역시 관찰되지 않았다. 요컨대, 사후확률을 계산하여 분석체계를 활성화시켰는지 여부와 상관없이 참가자들은 숫자정보가 확률(57.7%) 보다는 빈도(20.80 명 중의 12 명이나 2,080 명 중의 1,200 명)로 제시할 때 질병으로 사망할 위험을 더 크다고 판단하였다. 사후확률에 대한 위험판단에서도 숫자양식의 효과가 검증되었고, 빈도범위가 좁든지(20.80 명 중의 12 명)나 넓든지(2,080 명 중의 1,200 명) 위험판단은 다르지 않았다.

연구 1은 Slovic 등(2000)의 연구와 일치해서 숫자양식의 효과를 얻었다. 즉, 사람

들은 기저율과 사후확률 모두 확률이 아닌 빈도로 제시될 때 위험을 더 크게 판단하였다. 연구 1은 참가자들이 사후확률을 계산하면 숫자양식에 따른 위험판단 차이가 사라질 것으로 예측하였다. 그러나 숫자양식의 효과는 여전히 관찰되었다. 계산여부에 상관없이 숫자양식 효과가 발견된 점과 사후확률을 정확히 추정한 사람들에게서도 숫자양식의 효과가 발견되었다는 점은 Hsee와 Rottenstreich(2004)의 결과와는 달리, 위험판단에서 분석체계 이외의 과정이 개입되어 있음을 시사한다.

빈도범위 효과를 관찰한 Yamagishi(1997)의 연구와 달리 본 연구 1은 이 효과를 얻지 못했다. 그 이유로 몇 가지 가능성을 생각해볼 수 있다. 그가 사용한 자극은 천식(0.87% 또는 4.59%)에서부터 암(12.86% 또는 24.14%)에 이르는 여러 원인별 사망률이었다. 본 연구 1에서는 가상의 질병에 대한 위험을 판단하도록 하였고, 큰 위험 가능성(12%)만을 다루었다. 연구 2는 Yamagishi (1997)의 연구와 일관되지 않은 결과를 얻은 이유를 밝히려고 기저율과 사후확률의 크기를 고확률과 저확률로 나누고 위험판단에서 숫자양식 효과 및 빈도범위 효과를 검증하고자 하였다.

## 연구 2: 위험판단은 확률 크기와 숫자양식에 따라 다른가?

연구 1은 예측과 달리 빈도범위에 따른 위험판단의 차이를 관찰하지 못했다. 연구 2는 연구 1에서 얻은 결과가 다른 크기의 위험가능성에 대해서도 나타나는지를 검토하려고 실시되었다. 연구 2에서는 기저율이 높을 때(고-기저율)와 낮을 때(저-기저율), 그리고 사후확률이 높을 때(고-사후확률)와 낮을 때(저-사후확률) 위험판단을 비교하였다. 작은 확률의 사건에서만 비율-편향 효과가 관찰된다는 사실에 비추어 저확률 조건에서 빈도범위 효과가 관찰되어야 한다.

## 방 법

## 참가자

서울 소재 대학의 학생 355명이 참가하였고 소정의 보상을 받았다. 이들은 연구 1에 참가하지 않았다.

## 설계

숫자양식 변인의 효과가 확률의 크기에 따라 다른지 검증하기 위해 기저율과 사후확률에 대한 위험평정을 종속변인으로 하여 숫자양식(3) x 확률 크기(2)의 참가자간 설계를 실시하였다. 연구 2는 저-기저율과 고-기저율 조건 모두에서 숫자양식 변인의 효과를 검증하였고, 저-사후확률과 고-사후확률 조건 모두에서도 숫자양식 변인의 효과를 검증하였다.

## 자극 및 절차

### 기저율

자극 재료는 연구 1에서 사용된 자극과 같았으나 기저율이 낮은 것(저-기저율)과 높은 것(고-기저율)으로 구분하였다(표 1). 저-기저율 조건에서 확률은 2%, 좁은 빈도범위는 100명 중 2명, 넓은 빈도범위는 10,000명 중 200명으로 제시하였다(조건별 55명, 61명, 58명). 고-기저율 조건은 연구 1과 같았다. 즉 확률은 12%, 좁은 빈도범위는 100명 중 12명, 넓은 빈도범위는 10,000명 중 1,200명으로 제시하였다(조건별 60명, 61명, 60명).

### 사후확률

연구 1에서 계산이 사후확률 위험판단에 영향을 주지 않는다는 것이 확인되었으므로 연구 2에서는 연구 1의 비계산조건처럼 사후확률 값을 그대로 제시하였다. 자극 재료는 연구 1에서 사용된 자극과 구성이 같았으나 사후확률이 낮은 것(저-사후확률)과 높은 것(고-사후확률)으로 구분하였다. 저-사후확률 조건에서 확률은

6.4%, 좁은 빈도범위는 31.4명 중 2명, 넓은 빈도범위는 3,142명 중 200명으로 제시하였다. 연구 1과 마찬가지로 고-사후확률일 때 확률은 57.7%, 좁은 빈도범위는 20.8명 중 12명, 넓은 빈도범위는 2,080명 중 1,200명으로 제시하였다(표 1).

## 결과 및 논의

### 숫자양식과 기저울 크기에 따른 위험판단

숫자양식과 기저울 크기를 참가자간 변인으로 하여 이원변량분석을 실시하였다. 숫자양식의 주효과, 기저울 크기의 주효과, 그리고 숫자양식과 기저울 크기의 상호작용효과가 모두 통계적으로 유의하였다( $R(2, 349) = 16.30, MSe = 52.19, p = .000$ ;  $R(1, 349) = 20.81, MSe = 66.62, p = .000$ ;  $R(2, 349) = 4.12, MSe = 13.20, p = .017$ ) (그림 2).

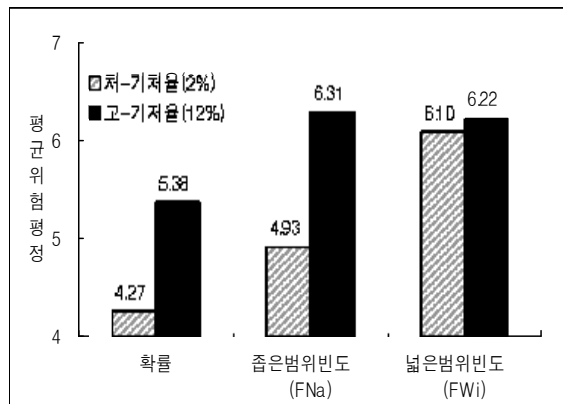


그림 2. 숫자양식과 기저울 크기별 평균 위험평정

숫자양식과 기저울 크기의 상호작용효과가 나타났으므로 기저울 크기별로 위험평정을 분석하였다. 계획된 사전비교에 의하면 고-기저울 수준에서 확률조건( $M = 5.38, SD = 1.77$ )보다 빈도조건( $M = 6.26, SD = 1.62$ )에서 더 위험하다고 판단되었

다( $t = 3.33$ ,  $df = 178$ ,  $p = .001$ ). 두 빈도조건 FNa 조건(100명 중 12명 사망)과 FWi 조건(10,000명 중 1,200명 사망)을 비교하였으나 통계적으로 차이가 없었다(FNa  $M = 6.31$ ; FWi  $M = 6.22$ ;  $t = .311$ ,  $df = 178$ ,  $p = .756$ ). 고-기저율 수준의 결과는 연구 1의 결과를 반복한다.

저-기저율 수준에서도 확률조건( $M = 4.27$ ,  $SD = 1.85$ )보다 빈도조건( $M = 5.50$ ,  $SD = 2.04$ )에서 더 위험하다고 판단되었다( $t = 4.02$ ,  $df = 171$ ,  $p = .000$ ). 두 빈도범위 조건 간에도 통계적으로 유의한 차이가 관찰되었다. 즉, 저-기저율 수준에서는 좁은 빈도범위(100명 중 2명 사망)보다 넓은 빈도범위(10,000명 중 200명 사망)일 때 질병 A가 훨씬 더 위험하다고 판단되었다(FNa  $M = 4.93$ ; FWi  $M = 6.10$ ,  $t = 3.35$ ,  $df = 171$ ,  $p = .001$ ).

### 숫자양식과 사후확률 크기에 따른 위험판단

숫자양식과 사후확률 크기를 참가자간 변인으로 하여 이원변량분석을 실시하였다. 사후확률 크기의 주효과와 숫자양식의 주효과는 통계적으로 유의하였으나( $F(1, 349) = 19.48$ ,  $MSe = 53.89$ ,  $p = .000$ ;  $F(2, 349) = 3.86$ ,  $MSe = 10.67$ ,  $p = .022$ ), 숫자양식과 사후확률 크기의 상호작용효과는 관찰되지 않았다( $F(2, 349) = .62$ ,  $MSe = 1.71$ ,  $p = .540$ ). 참가자들은 저-사후확률일 때보다 고-사후확률일 때 위험을 크게 판단하였다(저-사후확률  $M = 5.76$ ; 고-사후확률  $M = 6.25$ ). 계획된 사전비교에 의하면 확률조건( $M = 5.70$ ,  $SD = 1.80$ )보다 빈도조건( $M = 6.19$ ,  $SD = 1.66$ )에서 더 위험하다고 평정되었으며( $t = 2.56$ ,  $df = 352$ ,  $p = .011$ ), 두 빈도범위 조건에서는 위험판단의 차이가 없었다(FNa  $M = 6.25$ ; FWi  $M = 6.14$ ,  $t = .501$ ,  $df = 352$ ,  $p = .617$ ) (그림 3).

연구 2의 결과를 요약하면 다음과 같다. 확률의 크기에 상관없이 기저율과 사후확률 모두에서 숫자양식의 효과가 발견되었다. 그러나 빈도범위의 효과는 저-기저율에서만 관찰되었다. 다시 말해 질병으로 사망할 확률이 낮은 경우에만 넓은 범위 또는 좁은 범위에서 제시되었는지에 따라 위험판단이 달라졌다. 이 결과는 종합논의에서 다를 것이다.

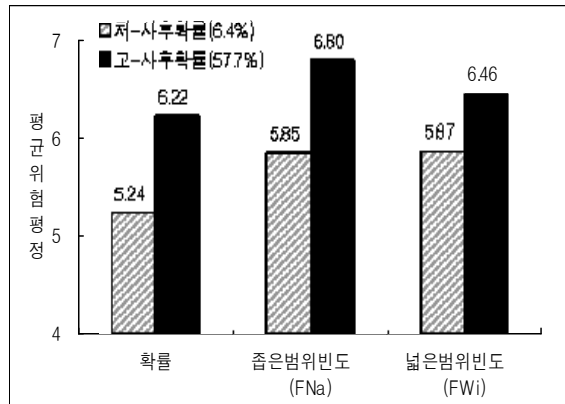


그림 3. 숫자양식과 사후확률 크기별 평균 위험평정

### 종합논의

본 연구는 베이스 추론 과제에서 위험에 관한 정보를 제시할 때 확률 또는 빈도양식이 위험판단과 추론 정확성에 미치는 영향을 검토하였다. 위험판단의 측면에서 보면, 기저율 정보를 확률이나 빈도로 제시할 때 판단이 달라졌다. 참가자들은 위험정보가 확률보다는 빈도로 제시되었을 때 질병에 의한 사망 가능성을 더 위험하다고 판단하였다(연구 1, 연구 2). 위험을 더 크게 판단하는 빈도의 효과는 기저율의 크기에 따라 달랐다(연구 2). 기저율이 작을 때만 빈도범위의 효과가 관찰되었다. 사후확률 정보에 대한 위험판단의 경우 숫자양식에 따라 위험판단이 달라졌으나 빈도범위의 효과는 관찰되지 않았다(연구 1). 계산과 같은 분석적 사고가 위험판단에 주는 영향을 검토하고자 계산조건과 비계산 조건을 비교하였으나 조건간의 차이가 발견되지 않았다. 추론의 정확성 측면에서 보았을 때, Gigerenzer와 Hoffrage(1995)가 발견한 숫자양식에 따른 차이가 한국 대학생 대상으로 반복되었다. 위험정보를 빈도로 제시하면 참가자들은 주어진 정보로부터 사후확률 즉 ‘양성평정을 받은 사람이 실제로 질병에 걸렸을 가능성’을 더 정확히 추정하였다(연구 1).

이중처리체계 접근은 본 연구의 결과들을 경험체계의 작용으로 설명한다. 질병



에 관한 내용의 자극은 감정을 동반한 심상을 유발한다. 숫자양식에 따라 그 정보가 불러일으키는 심상이 위험판단에 영향을 줄 가능성이 있다. 질병으로 인한 사망의 위험판단에서 ‘12%’라는 정보가 제시되면 이 때 떠오르는 심상은 한 사람이 되지 않으므로 위협하게 느껴지지 않는다. 그러나 ‘100명 중 12명’ 또는 ‘10,000명 중 1,200명’이 사망한다는 빈도조건에서는 다수의 사망자들의 심상이 떠올라 위험하다고 판단될 것이다.

앞서 소개된 빨간 구슬이 담긴 쟁반을 선택하는 경우처럼 사람들은 분모를 무시하는 심적 방략(Finucane, Peters와 Slovic, 2003)을 사용한다. 이 때 분모(기준이 되는 빈도범위)를 무시하고 분자인 사망자 수에 주목하여 위험판단을 내릴 것이다. 기저율에 대한 위험판단에서 빈도범위에 따라 사망자가 12명이나 1,200명으로 표현되므로 1,200명이 사망하는 조건에서 위험을 더 크게 판단할 것이다. 본 연구의 결과들은 이런 예측을 부정한다. 즉, 빈도범위에 따른 위험판단은 확률 크기에 따라 차이가 있었다. 비율 편향 효과는 작은 확률의 사건에 대해서만 관찰된다. 이 효과는 확률 크기에 따라 빈도범위 효과가 다른 관찰된 사실을 설명한다. 본 연구 2의 사망률 12%인 고-기저율의 경우, 두 빈도범위 조건에서 위험판단의 차이가 없었다. 사망자 12명이나 1,200명에서 모두 다수의 사망자란 심상을 불러일으키므로 둘 다 위험하다고 판단되었을 것이다. 반면, 사망률 2%인 저-기저율의 경우, 두 빈도범위 조건에서 위험판단의 차이가 나타났다. 분모를 무시하면 사망자 2명은 사망자가 거의 없다는 심상을 유발하여 위험하지 않다는 인상을 준다. 사망자 200명은 다수의 사망자 심상을 떠올리게 하므로 더 위험하다고 판단된다.

사후확률을 고려한 양성판정에 대한 위험판단에는 어떤 과정이 개입하였을까? 분모가 위험평정에 기여하는 정도가 기저율과 사후확률에서 다를 가능성이 있다. 기저율을 나타내려고 빈도조건에서 사용된 분모(100 명 또는 10,000 명)는 쉽게 무시되는 일반적 단위이다. 그러나 사후확률에서 사용된 분모(예, 20.8 명 또는 2,080 명)는 일반적인 단위가 아니어서 고려될만한 중요한 정보였을 것이다. 본 연구에서 관찰된 기저율과 사후확률에 대한 위험평정의 차이를 위와 같이 설명할 수 있다. 구체적으로 어떤 심리적 기제가 작동하였는지는 추후 연구에서 답을 찾아야 한다.

본 연구는 빈도정보를 제시할 때 범위의 효과가 항상 관찰되지 않음을 처음으로 밝혔다. 본 연구의 결과는 Yamagishi의 연구(1997)와 부분적으로 일치된다. Yamagishi

의 참가자들(1997)은 대다수의 사망원인에 대해 좁은 범위보다 넓은 범위에 제시된 빈도를 더 위험하다고 평정하였다. 본 연구의 참가자들은 사망률이 낮은 경우(2%)에 대해서만 빈도범위 효과를 보였다. 본 연구가 위험 확률의 크기를 연속선상에서 다루지 않아 빈도범위 효과가 어디서부터 나타나는지 알 수 없었다. 그러나 분명한 사실은 위험의 모든 기저율에서 빈도범위의 효과가 관찰되는 것은 아니라는 사실이다. 작은 사망률에 대해서만 빈도범위 효과가 관찰되었다는 점은 비율-편향 효과와 관련시켜 해석해 볼 수 있다. 다양한 확률 크기의 위험에 대한 판단을 다루면서, 특히 낮은 확률 크기의 위험을 판단할 때 사람들이 사용하는 심적 방략과 사고 체계의 영향을 밝혀야 한다.

본 연구의 결과들로 미루어 건강위험에 대해 숫자정보를 어떤 양식으로 제시하는가는 사소한 문제가 아니다. 확률로 제시된 정보는 이해하기 어려워 정확하게 소통이 되지 않는다. 예를 들어, 기상예보에서 흔히 언급되는 ‘비 올 확률 30%’의 정확한 의미를 잘 모르는 사람들도 많다(Gigerenzer & Edwards, 2003). 건강위험에 관해 의학적인 결정을 내릴 때 사람들은 숫자정보에 근거해서 결정을 한다. 위험정보를 확률로 제시하면 이해하기 어려울 뿐 아니라 위험판단도 영향을 받는다. 의학적 결정에 관한 베이스 추론과 위험판단에 대한 본 연구는 일반인들이 위험정보를 이해하고 결정할 때 분석적으로 사고하도록 유도하는 것만으로는 부족함을 보여주었다. 사후확률을 정확하게 추정한 사람들조차 확률과 빈도 양식에서 차이 나는 위험판단을 보였다는 사실은 베이스 추론 과제에서 정확한 추론 이외에 다른 기제가 위험판단에 작용하고 있음을 시사한다.

## 참고문헌

- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, 2, 206-214.
- Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing:

- when people behave against their better judgment. *Journal of personality and social psychology*, 66, 819-829
- Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, 21, 1083-1092.
- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities, In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*(pp.249-267). Cambridge, England: Cambridge University Press.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49, 709-724.
- Finucane, M. L., Peters, E., & Slovic, P. (2003). Judgment and decision making: The dance of affect and reason. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research*. UK: Cambridge University Press.
- Gigerenzer, G. (1996). The psychology of good judgment: frequency formats and simple algorithms. *Journal of Medical decision making*, 16, 273-280.
- Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal*, 327, 741-744.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, 102, 684-704.
- Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, 73, 538-540.
- Hsee, C., & Rottenstreich, Y. (2000). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General*, 1, 23-30.
- Kahneman, D., & Frederick. S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics of intuitive judgment: Extensions and Applications*. New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.
- Kirkpatrick, L., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of Personality and Social*

- Psychology*, 63, 534-544.
- Krynski, T., & Tenenbaum, J. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, 136, 430-450.
- Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76, 972-987.
- Slooman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- Slovic, P., & Monahan, J. (1995). Probability, danger and coercion: a study of risk perception and decision-making in mental health law. *Law and Human Behavior*, 19, 49-65.
- Slovic, P., Monahan, J., MacGregor, D. (2000). Violence risk assessment and risk communication: the effects of using actual cases providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, 24, 271-296.
- Yamagishi, K. (1997). When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. *Applied Cognitive Psychology*, 11, 495-506.
- Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: the role of representation in mental computation. *Cognition*, 98, 287-308.

1 차원고접수 : 2009. 8. 11

2 차원고접수 : 2009. 9. 21

최종게재승인 : 2009. 9. 24

(*Abstract*)

## Effects of Numerical Formats and Frequency ranges on Judgment of Risk and Inference in the Bayesian Inference Task

Hyunju Lee

Young-Ai Lee

Department of Psychology, Ewha Women's University

We examined risk judgment and the accuracy of inference based on two kinds of probabilities in a Bayesian inference task: the death probability from a disease (base rates) and the probability of having a disease with positive results in the screening test (posterior probabilities). Risk information were presented in either a probability or a frequency format. In Study 1, we found a numerical format effect for both base rate and posterior probability. Participants rated information as riskier and inferred more accurately in the frequency condition than in the probability condition for both base rate and posterior probability. However, there was no frequency range effect, which suggested that the ranges of frequency format did not influence risk ratings. In order to find out how the analytic thought system influences risk ratings, we compared the ratings of a computation condition and those of a no-computation condition and still found the numerical format effect in computation condition. In Study 2, we examined the numerical format effect and frequency range effect in a high and a low probability condition and found the numerical format effect at each probability level. This result suggests that people feel riskier in the frequency format than in the probability format regardless of the base rates and the posterior probability. We also found a frequency range effect only for the low base rate condition. Our results were discussed in terms of the dual process theories.

*Keywords* : Numerical information format, frequency range, probability, frequency, Bayesian inference, risk judgment