

# Analysis and Estimation for Market Share of Biologics based on Google Trends Big Data

Ki Tae Bong · Heesang Lee<sup>†</sup>

Department of Management of Technology, Sungkyunkwan University

## 구글 트렌드 빅데이터를 통한 바이오의약품의 시장 점유율 분석과 추정

봉기태 · 이희상<sup>†</sup>

성균관대학교 기술경영전문대학원

Google Trends is a useful tool not only for setting search periods, but also for providing search volume to specific countries, regions, and cities. Extant research showed that the big data from Google Trends could be used for an on-line market analysis of opinion sensitive products instead of an on-site survey. This study investigated the market share of tumor necrosis factor-alpha (TNF- $\alpha$ ) inhibitor, which is in a great demand pharmaceutical product, based on big data analysis provided by Google Trends. In this case study, the consumer interest data from Google Trends were compared to the actual product sales of Top 3 TNF- $\alpha$  inhibitors (Enbrel, Remicade, and Humira). A correlation analysis and relative gap were analyzed by statistical analysis between sales-based market share and interest-based market share. Besides, in the country-specific analysis, three major countries (USA, Germany, and France) were selected for market share analysis for Top 3 TNF- $\alpha$  inhibitors. As a result, significant correlation and similarity were identified by data analysis. In the case of Remicade's biosimilars, the consumer interest in two biosimilar products (Inflectra and Renflexis) increased after the FDA approval. The analytical data showed that Google Trends is a powerful tool for market share estimation for biosimilars. This study is the first investigation in market share analysis for pharmaceutical products using Google Trends big data, and it shows that global and regional market share analysis and estimation are applicable for the interest-sensitive products.

**Keywords** : Consumer Interest, Google Trends, Market Share, TNF- $\alpha$  Inhibitor, Biosimilar

### 1. 서 론

최근 기업들은 제품에 대한 소비자의 관심도(Interest)를 높이기 위해 트위터(Twitter), 페이스북(Facebook), 유튜브(YouTube)와 같은 소셜 네트워크 시스템(Social Network System, SNS)을 활용하고 있다[19]. 기업들은 SNS상의

소비자들의 관심도가 자사의 제품 검색으로 나타나므로 이를 분석하여 제품개발이나 판매에 효과적으로 활용하고자 한다.

최근에는 전 세계적으로 가장 널리 쓰이는 구글 검색 엔진의 검색량이 '구글 트렌드(Google Trends)'라는 빅데이터 분석도구를 통해 시·공간적으로 구분하여 수집되어 대중적 관심도 수치로 나타낼 수 있고 체계적으로 분석할 수 있게 되었다[7]. 구글 트렌드는 대상 지역의 대상 기간 중 검색 횟수가 가장 많았던 때를 100으로 정하고, 다른 기간의 검색량을 0~100 척도(scale)로 하여 표준화한

관심도 수치를 제공하는 빅데이터 도구이다. 구글 트렌드는 특정 기간 및 국가, 지역, 도시 단위까지 관심도 수치를 제공하는 측면에서 매우 유용한 빅데이터 분석도구라 할 수 있고, 마케팅 및 고객 서비스 등 기업 경영의 활동을 최적화하고 파악하는데도 도움을 제공하고 있다[17]. 구글 트렌드는 최대 5개까지 키워드를 입력하여 비교할 수 있어 독과점 형태의 경쟁 제품 간 관심도 비교에 적합하게 사용될 수 있다.

바이오 제약 산업은 의사와 환자라는 2가지 계층의 소비자를 갖는 점이 일반 제품과는 다르다. 즉, 처방전이 필요한 의약품의 경우에는 최종소비자인 환자와 구매 결정권자인 의사라는 결정소비자가 동시에 존재하고 이 두 집단의 관심도가 상호작용하여 최종 구매 결정이 이루어지는 특수한 상황이 발생한다. 또한 약가, 품목허가기준, 광고기준 등 국가별 규제사항도 일반 제품과는 달리 엄격하므로 시장 조사 및 SNS 활용 등에 있어 타 산업과는 다른 고유의 특성을 지니고 있다.

바이오 제약 산업에서 최근 빠르게 발전하고 있는 바이오의약품 시장 및 바이오시밀러 시장이 주목 받고 있다. 생명공학 정책연구센터가 분석한 결과에 따르면 전 세계 바이오시밀러 시장은 2019년 241억 달러(약 29조 원) 규모에서 2023년 481억 달러(약 57조 원) 규모로 확대될 전망이다[4]. 바이오의약품 중 특히 중앙괴사인자-알파(Tumor Necrosis Factor-alpha, TNF- $\alpha$ ) 저해제 시장은 전체 제약품 시장 중 가장 큰 시장을 형성하고 있고, 엔브렐(Enbrel), 레미케이드(Remicade), 휴미라(Humira) 등 글로벌 상위 3개 제품이 독과점하고 있다. 이들 상위 3개 제품은 류마티스 관절염 같은 자가면역 질환의 치료에 사용되며, 2016년 글로벌 판매량이 휴미라 161억 달러, 엔브렐 89억 달러, 레미케이드 78억 달러를 보여 전 세계 모든 제약품 중 각각 1위, 3위, 5위의 매출 실적을 올렸다[2]. 현재, 엔브렐, 레미케이드, 휴미라 등 3개 제품에 대한 생물학적 복제약인 바이오시밀러의 개발도 활발히 이루어지고 있으며, 노바티스(Novartis), 화이자(Pfizer) 등 글로벌 제약사와 함께 우리나라 기업인 셀트리온(Celltrion), 삼성바이오에피스(Samsung Bioepis) 등이 참여하는 TNF- $\alpha$  저해제의 바이오시밀러 시장이 형성되고 있다.

본 연구는 구글 트렌드 빅데이터를 활용하여 글로벌 상위 3개 TNF- $\alpha$  저해제의 소비자 관심도 점유율과 매출액 점유율을 분석하고 이를 바탕으로 우리나라가 강점을 보이고 있는 바이오시밀러 제품의 성장 가능성을 추정하고자 한다. 본 연구의 제 2장에서는 구글 트렌드 및 의약품의 시장 점유율 관련한 기존의 연구를 고찰하였고, 구글 트렌드를 통한 바이오의약품의 시장 점유율 분석의 연구 모델 및 연구 방법을 제시하였다. 제 3장에서는 실제 매출액과 구글 트렌드에 나타난 소비자 관심도 수치의 비

교를 통해 글로벌 상위 3개의 TNF- $\alpha$  저해제 제품인 엔브렐, 레미케이드, 휴미라의 관심도 점유율과 매출액 점유율을 분석하였고, 주요 의약품 시장인 글로벌 3개국(미국, 독일, 프랑스)에서의 시장 점유율도 분석하였다. 특히, 레미케이드 오리지널 제품의 시장을 잠식하고자 국내 제약사인 삼성바이오에피스와 셀트리온에서 미국 시장에 출시한 2개의 바이오시밀러 제품에 대해 구글 트렌드 기반의 관심도 점유율 분석을 통해 매출액 점유율을 추정하였다. 끝으로, 4장에서는 본 연구가 갖는 학문적 기여와 시사점을 기술하였다.

## 2. 이론적 배경 및 방법

### 2.1 기존의 연구

구글 트렌드 빅데이터 분석을 질병 예측에 사용하였던 대표적인 연구는 Dugas et al.[8]이 수행한 구글 트렌드 데이터와 독감환자의 증감의 상관관계에 대한 연구인데, 일반인들의 관심도 데이터를 사용해 독감환자의 발생률에 대한 예측과 실시간으로 필요한 치료 백신의 양을 예측하였다. 또한, 국가 정책의 수립에 있어서는 나노 기술에 대한 키워드를 기반으로 구글 트렌드 빅데이터를 활용하여 나노 기술 트렌드를 예측하고 이를 전략적으로 활용한 연구[18]가 있었다. 기업 활동과 관련한 구글 트렌드의 데이터 활용 가능성을 보여준 연구로는 비엠더블유(BMW)와 메르세데스 벤츠(Mercedes-Benz) 사를 중심으로 소비자들의 구글 트렌드 검색량이 제품의 매출액 및 제조사의 주가와 양의 상관관계를 보임을 증명한 연구[22], 대형 소매업체에서 매출액과 구글 트렌드의 소비자 관심도 데이터 간 유의한 연관성이 있음을 보여준 연구[6], 구글 트렌드에서 TV와 관련된 검색어를 통해 국가마다 빈도수를 측정하여 TV 시장을 예측하는데 도움이 된다는 사실을 밝힌 연구[21] 등이 있다. 지금까지의 구글 트렌드 빅데이터를 활용한 연구들은 독감, 하이테크 기술/제품(나노 기술, 고급승용차, TV 등)을 대상으로 하는 연구들이었고, 이와 같은 질병, 제품, 기술의 확산에는 사람들의 관심도가 큰 영향을 미친다는 공통점이 있었다.

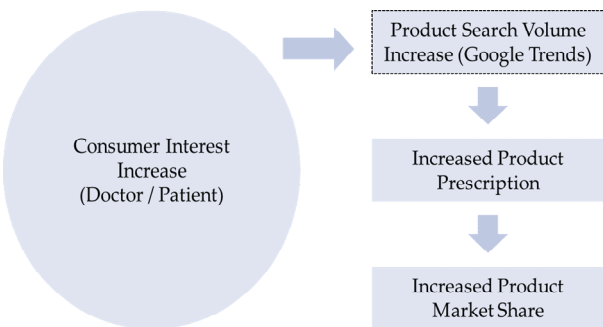
의약품의 시장 점유율과 관련한 연구는 주로 매출액의 시장 점유율에 미치는 요인, 그리고 시장 점유율에 대한 추정 등에 관한 연구가 이루어졌다. 시장 점유율에 미치는 요인에 관한 연구로 Magazzini et al.[14]은 미국 및 유럽 3국 의약품 시장에서 특히 만료 후 의약품들의 매출액 점유율이 제품 가격과 관계가 있음을 회귀분석 방법으로 분석하였고, 배은영[1]은 1998년 상반기 매출 자료를 통하여 시장 진입 순위가 복제약의 시장 점유율 결정요인으로서

가장 큰 영향력을 가진 것으로 분석하였다. 오근엽[16]은 복제약 시장의 매출액 점유율을 로지스틱 함수를 사용하여 추정하고 실제 매출액 점유율에 매우 근접한 로지스틱 함수를 얻을 수 있었다. 하지만, 현재까지 이루어진 연구에서는 의약품의 시장 점유율 파악에 있어서 소비자의 관심도를 반영하거나 구글 트렌드의 빅데이터를 활용한 연구는 없었다. 그러나 바이오의약품 시장은 독감처럼 사회적 관심이 큰 중요 질병에 깊은 관련이 있고, 글로벌 시장을 과점한 몇 가지 혁신적인 제품간 경쟁이 치열하여 소비자의 관심도가 매출액 점유율에 잘 반영되는 상품이라는 것을 감안한다면 구글 트렌드를 이용한 소비자 관심도 분석을 통해 시장의 매출액 점유율을 분석하는 것은 매우 효과적일 것이다.

## 2.2 연구 방법

### 2.2.1 연구모델 설정

의약품은 주요 질병에 대한 환자들의 우려와 의사들의 처방이 동반된다. 의사와 환자 모두 새로운 의약품의 개발과 출시에 관심이 높고, 매출은 의사의 처방에 따라 결정된다. 본 연구에서는 <Figure 1>에서 보는 바와 같이 구글 검색으로 누적된 검색량의 빅데이터가 실제 시장의 매출액 점유율로 나타나는지를 검증하는 연구 모델을 설정하였다. 이 모델은 의사와 환자 두 집단으로 구성된 소비자의 관심도 증가가 결국 바이오의약품에 대한 처방 증가로 이어져 시장의 매출을 향상시킬 것이라는 가정에 기초한다. 본 연구에서는 이러한 연구모형에 따라 실제 매출액 점유율과 소비자 관심도 점유율과의 상관 분석 및 유사도 분석을 시계열적으로 수행하여 이를 입증하고자 하였다.



<Figure 1> The Model of Product Market Share Increase in Biopharmaceutical Industry

### 2.2.2 제품 데이터 및 빅데이터 수집

제품 데이터의 구글 트렌드 분석에 사용할 검색어와 검색 기간을 설정하기 위하여 미국식품의약국(Food and Drug

Administration, FDA)에서 제공하는 퍼플북(Purple Book) 정보[9]를 다운 받아 사용하였다. 이를 바탕으로 구글 트렌드 분석의 검색 키워드는 오리지널 제약사의 글로벌 상위 3개 제품인 TNF- $\alpha$  저해제의 영문명(Enbrel, Remicade, Humira)을 선정하였고, 레미케이드 제품의 바이오시밀러 2 제품의 경우도 영문명(Inflectra, Renflexis)을 키워드로 선정하였다. 검색 기간 설정은 오리지널 제품의 판매 허가 년도, 바이오시밀러 제품의 판매 허가 년도를 확인 후 적용하였으며, 구글 트렌드를 통하여 추출한 월별 관심도 데이터 수치를 누적하여 연간 실제 매출액과 비교하였다.

본 연구에서는 오리지널 TNF- $\alpha$  저해제인 엔브렐, 레미케이드, 휴미라의 시장 매출액 점유율은 바이오 의약 특허 정보 시스템[11]에 보고된 2006년~2014년 9년 동안의 글로벌 매출액 정보를 수집하였고, 비교를 위해 동일 기간의 구글 트렌드 관심도 데이터를 수집하였다. 지역별 시장 매출액 점유율은 주요 판매 국가인 미국, 독일, 프랑스에서의 2013년~2015년 3년 동안의 품목별 매출액[5]과 해당 기간 구글 트렌드에서 나타나는 지역별 소비자 관심도 수치 데이터를 수집하여 비교하였다. 특히, 미국시장에서는 오리지널 제품인 레미케이드에 대해 셀트리온이 2016년도에 바이오시밀러 제품인 인플렉트라(Inflectra)를 출시하였고, 삼성바이오에피스가 2017년 렌플렉시스(Renflexis)를 출시하였기 때문에, 2015년~2017년 3년 동안의 관심도 데이터를 구글 트렌드로 수집하여 매출액 점유율 추정 모형을 만들었다. 추정된 모형의 정확도 검증을 위해 2018년도 인플렉트라와 실제 매출액 자료[3]와 렌플렉시스의 매출액 추정 자료[12]를 수집하였다.

### 2.2.3 분석 과정 및 방법

구글 트렌드 빅데이터를 통해 얻어진 소비자의 관심도 점유율이 바이오의약품의 매출액 점유율과 밀접한 관계를 갖는지를 상관 분석하였다. 상관 분석은 변수와 변수 사이의 관계를 이해할 수 있고, 선형성의 정도에 대한 분석 결과를 제공하기 때문에 분석 방법으로 채택되었다. 데이터는 엑셀 파일 형태로 추출하여 저장한 후 기간별로 정리하였고, 미니탭 19(Minitab 19) 프로그램을 이용하여 상관 계수의 강도를 계산하였다.

상관 분석과 더불어 매출액 점유율과 구글 트렌드에 나타난 관심도 점유율 간의 유사 정도를 판단하기 위하여 상대 격차(Relative gap) 지표를 도입하였다. 상대 격차는 아래와 같은 수식으로 계산하였으며, 이는 매출액 점유율과 관심도 점유율의 평균값에 대한 상대적인 차이를 나타내는 유사도 수치이다. 상대 격차는 두 시장 점유율 값 사이의 유사도를 -2.00~2.00의 수치로 나타낼 수 있으며, 0.00에 가까울수록 격차(gap)가 작아져 두 데이터가 일치함을 보여준다.

<Table 1> Strength of Correlation

Range	Strength of Correlation
0.01~0.30	Very Weak
0.31~0.50	Weak
0.51~0.70	Moderate
0.71~0.90	Strong
0.91~1.00	Very Strong

<Table 2> Strength of Relative Gap

Range	Strength of Relative Gap
-0.20~0.20	Very Low
-0.40~-0.21, 0.21~0.40	Low
-0.60~-0.41, 0.41~0.60	Moderate
-0.80~-0.61, 0.61~0.80	High
-2.00~-0.80, 0.80~2.00	Very High

<Table 1>에서는 상관 계수의 강도를 판단하는 기준을 보여주며[20], <Table 2>에서는 0.2 구간 단위로 구분하여 상대 격차의 강도를 나타내었다.

- 상대 격차 수식

$$Gap\ ratio(t) = \frac{X_t - Y_t}{\left(\frac{X_t + Y_t}{2}\right)}$$

- $X_t = t$  시점의 관심도(interest) 기반의 시장 점유율 (market share)
- $Y_t = t$  시점의 매출(sales) 기반의 시장 점유율(market share)

상관 계수는 관심도와 매출액이 얼마나 비례 관계가 있는지를 측정하는 반면, 상대 격차는 관심도 및 매출액 점유율 차이가 매출액대비 얼마나 큰 격차를 보이는지 계산할 수 있다. 특히 양수(+)인 경우 매출액 점유율보다 관심도 점유율이 낙관적임을 보여주고, 음수(-)인 경우 매출액 점유율보다 관심도가 비관적임을 보여주기 때문에 <Figure 1>의 연구모델이 가정했던 관심도 점유율 변화가 매출액 점유율 변화에 선행하는지를 검증하는 지표로 활용이 가능하다. 즉, 소비자의 관심도가 증가하면 네트워크 효과로 기인한 제품 구매 결정 요인으로 작용하여 매출이 증대될 것이라는 기존 이론[13]을 검증할 수 있다.

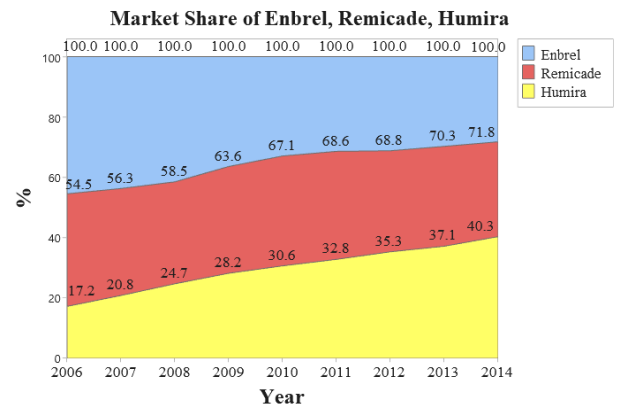
### 3. 연구 결과

#### 3.1 바이오의약품의 시장 점유율 분석

오리지널 제약사의 글로벌 매출 상위 3개의 TNF- $\alpha$  저해제 제품인 엔브렐, 레미케이드, 휴미라에 대하여 2006

<Table 3> Annual Sales Change of Top 3 TNF- $\alpha$  Inhibitors (Million Dollars)

Year	Enbrel Sales (M USD)	Remicade Sales (M USD)	Humira Sales (M USD)
2006	5,396	4,429	2,044
2007	6,434	5,218	3,064
2008	7,650	6,228	4,552
2009	7,215	7,010	5,583
2010	7,245	8,012	6,737
2011	7,886	8,991	8,233
2012	8,514	9,136	9,628
2013	8,894	9,900	11,105
2014	9,120	10,151	13,021

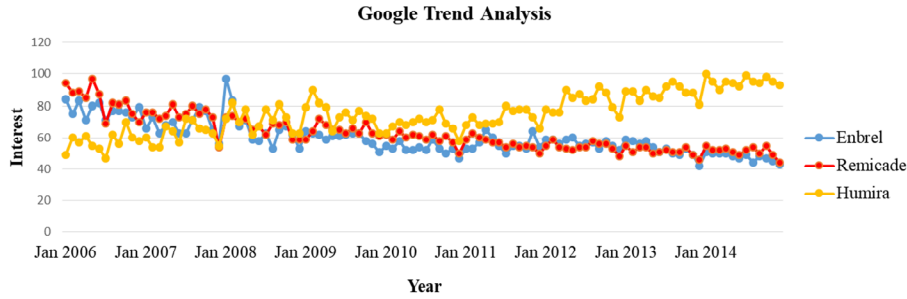


<Figure 2> Sales based Market Share Change of Top 3 TNF- $\alpha$  Inhibitors

년부터 2014년까지 발생한 품목별 매출액을 <Table 3>에 나타내었다[11].

<Figure 2>는 <Table 3>의 자료로부터 2006년부터 2014년까지의 글로벌 상위 3개의 TNF- $\alpha$  저해제의 매출액 점유율 변화를 도식화한 결과이다. 휴미라의 경우 2006년 3제품 중 가장 낮은 17.2%의 매출액 점유율을 보이다가 2014년 가장 높은 40.3%의 점유율을 보이고 있다. 휴미라의 경우 1998년 미국식품의약국 허가를 받은 레미케이드와 엔브렐보다 4년이 늦은 2002년에 미국식품의약국 허가를 받은 후발 제품이긴 하지만, 우수한 약효와 기술적 진보 그리고 공격적인 마케팅으로 시장 지배력을 증가시켰음을 알 수 있다[10].

<Figure 3>은 구글 트렌드에 나타난 글로벌 상위 3개의 TNF- $\alpha$  저해제의 2006년~2014년 동안의 소비자 관심도 변화이다. 제품별 매출액에서 나타난 점유율의 경향성은 구글 트렌드 분석의 관심도 점유율에서도 유사하게 나타났다. 즉, 매출액을 통한 분석과 마찬가지로 휴미라가 관심도 점유율 2006년 3등에서 출발하여 2010년 이후 1등으로 등극하고, 레미케이드와 엔브렐의 경우는 2010년 이후 관심도 점유율에서도 2등을 다투는 경쟁 구도를 보이는 것을 확인할 수 있다.



<Figure 3> Consumer Interest based Market Share Change of Top 3 TNF-α Inhibitors

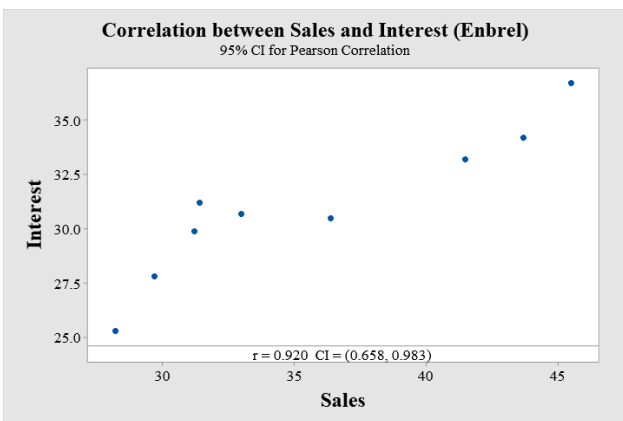
매출 규모에 따른 시장 점유율과 소비자 관심도에 따른 시장 점유율의 관계를 통계적으로 분석하기 위해 피어슨 상관 분석을 실시하였다. <Figure 4>는 엔브렐에 대한 상관 분석 결과를 보여주며, 매우 강한 양의 상관관계(피어슨 상관계수 0.920)가 유의성 있게(p-value 0.000) 나타났다. <Table 4>에 나타난 상대 격차의 경우 평균 -0.13으로 매우 낮아, 관심도 점유율과 매출액 점유율 사이에 높은 유사도를 보이는 것을 알 수 있다. 또한, 상대격차가 9년간 지속적으로 음수(-)여서 엔브렐의 경우 관심도 점유율이 매출액 점유율보다 지속적으로 비관적임을 알 수 있다. 엔브렐의 실제 매출액 점유율은 시간이 지남에 따라 점점 감소하여 관심도 점유율 변화가 선행함을 잘 보여주었다.

올보다 낙관적이다가 곧 음수(-)로 바뀌어 비관적으로 전환됨을 알 수 있다. 레미케이드의 실제 매출액 점유율은 엔브렐처럼 시간이 지남에 따라 점점 감소하여 관심도 점유율 변화가 선행함을 잘 보여주었다.

<Table 4> Market Share of Enbrel Product among Top 3 TNF-α Inhibitors (Sales vs Interest)

Year	Enbrel (Interest), %	Enbrel (Sales), %	Simple Difference	Relative Gap
2006	36.7	45.5	-8.8	-0.21
2007	34.2	43.7	-9.5	-0.24
2008	33.2	41.5	-8.3	-0.22
2009	30.5	36.4	-5.9	-0.18
2010	30.7	33.0	-2.3	-0.07
2011	31.2	31.4	-0.2	-0.01
2012	29.9	31.2	-1.3	-0.04
2013	27.8	29.7	-1.9	-0.07
2014	25.3	28.2	-2.9	-0.11
Average			-4.6	-0.13

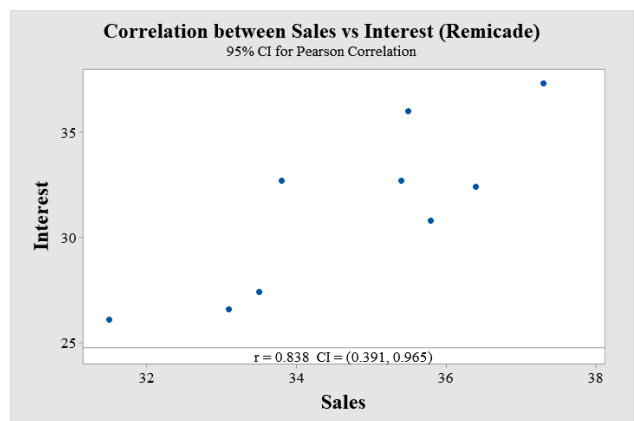
<Figure 5>는 레미케이드에 대한 상관 분석 결과를 보여주며, 엔브렐과 마찬가지로 강한 양의 상관 관계(피어슨 상관계수 0.838)가 유의성 있게(p-value 0.005) 나타났다. <Table 5>에 나타난 상대 격차의 경우도 평균 -0.11으로 매우 낮아, 관심도 점유율과 매출액 점유율 사이에 높은 유사도를 보여주었다. 상대 격차의 부호를 보면 초반은 양수(+)로 미세하게 관심도 점유율이 매출액 점유



Pairwise Pearson Correlations

Sample 1	Sample 2	N	Correlation	95% CI for ρ	P-Value
Interest	Sales	9	0.920	(0.658, 0.983)	0.000

<Figure 4> Correlation Analysis of Enbrel(Market Share of Sales vs Interest)



Pairwise Pearson Correlations

Sample 1	Sample 2	N	Correlation	95% CI for ρ	P-Value
Interest	Sales	9	0.838	(0.391, 0.965)	0.005

<Figure 5> Correlation Analysis of Remicade(Market Share of Sales vs Interest)

<Table 5> Market Share of Remicade Product among Top 3 TNF-α Inhibitors(Sales vs Interest)

Year	Remicade (Interest), %	Remicade (Sales), %	Simple Difference	Relative Gap
2006	37.3	37.3	0.0	0.00
2007	36.0	35.5	0.5	0.01
2008	32.7	33.8	-1.1	-0.03
2009	32.7	35.4	-2.7	-0.08
2010	32.4	36.4	-4.0	-0.12
2011	30.8	35.8	-5.0	-0.15
2012	27.4	33.5	-6.1	-0.20
2013	26.6	33.1	-6.5	-0.22
2014	26.1	31.5	-5.4	-0.19
Average			-3.4	-0.11

<Table 6> Market Share of Humira Product among Top 3 TNF-α Inhibitors (Sales vs Interest)

Year	Humira (Interest), %	Humira (Sales), %	Simple Difference	Relative Gap
2006	26.0	17.2	8.8	0.41
2007	29.8	20.8	9.0	0.36
2008	34.1	24.7	9.4	0.32
2009	36.8	28.2	8.6	0.26
2010	36.9	30.6	6.3	0.19
2011	38.0	32.8	5.2	0.15
2012	42.7	35.3	7.4	0.19
2013	45.6	37.1	8.5	0.21
2014	48.5	40.3	8.2	0.18
Average			7.9	0.25

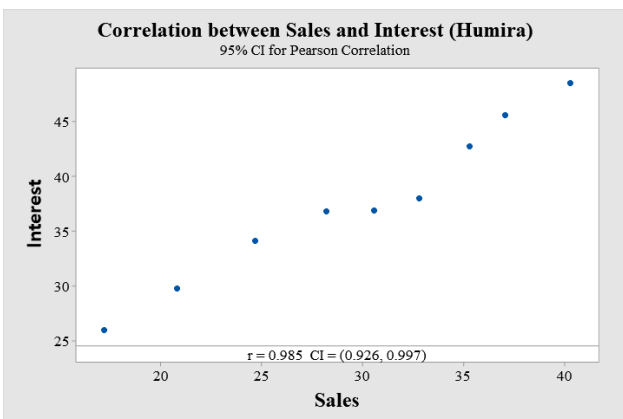
<Figure 6>에서 보는 바와 같이 휴미라의 경우 상관 분석 결과 매출액 점유율과 관심도 점유율 사이에 3개 경쟁 제품 중 가장 강한 양의 상관관계(피어슨 상관계수 0.985)가 유의성 있게(p-value 0.000) 나타났다. <Table 6>에 나타난 상대 격차의 경우는 평균 0.25로 낮아, 두 시장 점유율 수치 사이에 유사함을 보였다. 주목할 점으로 3가지 오리지널 제품 중 휴미라는 9년 동안 줄곧 상대 격차부호가 양수(+)만 나타나 소비자 관심도가 매출액 점유율보다 줄곧 낙관적인 경향을 나타냈고 실제 매출액 점유율은 시간이 지남에 따라 점점 증가하여 관심도 점유율 변화가 선행함을 잘 보여주었다.

글로벌 상위 3개의 TNF-α 저해제 바이오의약품의 시장 점유율을 분석한 결과, 구글 트렌드의 관심도 데이터 만으로도 강한 양의 상관관계와 높은 유사도로 제품별

매출액 점유율 분석이 가능함을 알 수 있었다. 상대 격차에 따른 매출액 점유율 추정에서 엔브렐은 줄곧 음수(-)값을 보였고, 레미케이드는 양수(+)값을 보이다가 이후 음수(-)값으로 바뀌었다. 반면, 후발주자이며 성능이 우수한 휴미라는 계속적으로 양수(+)값을 보이더니 결국 매출액 점유율 1등을 차지하였다. 이는 구글 트렌드 상에서 소비자의 관심도 증가가 나타나면, 직접적인 제품 매출액 증가로 이어지고 시장의 매출액 점유율이 상승함을 보여준다.

### 3.2 글로벌 시장의 주요 3국 시장 점유율 분석

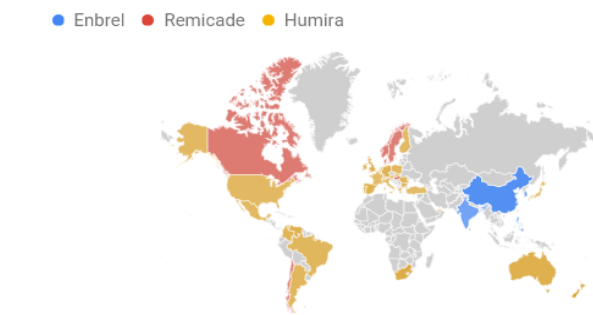
바이오의약품의 국가별 시장 점유율 분석을 위해 2013년~2015년 기간 동안 전 세계의 지역별 구글 트렌드 데이터를 분석하여 <Figure 7>을 얻을 수 있었다. 이를 살펴보면 글로벌 상위 3개의 TNF-α 저해제의 국가별 소비자의 관심도 분포를 알 수 있었는데, 엔브렐의 경우 중국과 대한민국, 레미케이드의 경우 덴마크와 캐나다, 휴미라의 경우 뉴질랜드와 불가리아에서 각각 관심도 점유율이 높음을 식별할 수 있다.



#### Pairwise Pearson Correlations

Sample 1	Sample 2	N	Correlation	95% CI for p	P-Value
Interest	Sales	9	0.985	(0.926, 0.997)	0.000

<Figure 6> Correlation Analysis of Humira(Market Share of Sales vs Interest)



<Figure 7> Regional Interests Analysis of Top 3 TNF-α Inhibitors

지역별 시장의 매출액 점유율 분석은 특정 지역에서 제품의 경쟁력과 시장 전망을 파악하기 위해 필요하다. 개별 국가에 대한 제품별 점유율을 살펴보기 위하여 바이오의약품 시장규모가 가장 큰 미국(2015년 기준 1,125억 달러)과 유럽 지역(2015년 334억 달러) 중 주요 시장인 독일, 프랑스에 대하여 제품별 시장 점유율 분석을 해보았다. 이들 3개국 시장은 글로벌 제약사들이 바이오의약품 출시에 있어서 가장 먼저 진출하는 시장이기도 한데, 품목별 매출액 점유율은 바이오의약품 산업분석보고서[5]를 통해 2013년~2015년 3년간 매출 데이터를 파악하고, 이를 구글 트렌드 기반의 소비자 관심도 점유율과 비교하였다. 상관 분석의 경우 입수할 수 있는 국가별 매출액 점유율 데이터가 3년치에 불과해 상관도 계수 계산은 수행하지 않았다.

<Table 7>은 미국 시장에서의 엔브렐, 레미케이드, 휴미라의 관심도와 매출액의 점유율을 보여준다. 2013년~2015년 3년간의 상대 격차 평균이 각각 -0.24, 0.06, 0.12의 낮은 수치를 보여, 매출액 점유율과 소비자 관심도 점유율은 제품별로 높은 유사도를 보이고 있음을 알 수 있다.

<Table 8>은 독일 시장에서의 엔브렐, 레미케이드, 휴미라의 관심도와 매출액의 점유율을 보여준다. 미국 시장과 마찬가지로 2013년~2015년 3년간의 상대 격차 평균이 각각 -0.28, 0.20, 0.06의 낮은 수치를 보여, 매출액 점유율과 소비자 관심도 점유율은 제품별로 높은 유사도를 보이고 있음을 알 수 있다.

<Table 9>는 프랑스 시장에서의 엔브렐, 레미케이드, 휴미라의 관심도와 매출액의 점유율을 보여준다. 2013년~2015년 3년간의 상대 격차 평균이 각각 -0.09, 0.01, 0.04의 낮은 수치를 보여, 매출액 점유율과 소비자 관심도 점유율은 높은 유사도를 보이고 있음을 알 수 있다.

<Table 7> Market Share in USA

Year	Enbrel (Interest), %	Enbrel (Sales), %	Simple Difference	Relative Gap
2013	27.6	32.9	-5.30	-0.18
2014	24.7	32.2	-7.50	-0.26
2015	22.6	29.7	-7.10	-0.27
Average			-6.63	-0.24
Year	Remicade (Interest), %	Remicade (Sales), %	Simple Difference	Relative Gap
2013	27.1	28.2	-1.10	-0.04
2014	27.0	25.0	2.00	0.08
2015	26.0	22.6	3.40	0.14
Average			1.43	0.06
Year	Humira (Interest), %	Humira (Sales), %	Simple Difference	Relative Gap
2013	45.4	38.9	6.50	0.15
2014	48.2	42.8	5.40	0.12
2015	51.4	47.8	3.60	0.07
Average			5.17	0.12

은 높은 유사도를 보이고 있음을 알 수 있다.

이상 주요 3개국 데이터에서 글로벌 상위 3개의 TNF-α 저해제의 매출에 따른 매출액 점유율과 구글 트렌드에서 나타나는 소비자 관심도 점유율은 높은 유사성을 보임을 알 수 있었다. 지역별 분석에서도 글로벌 분석에서와 마찬가지로 휴미라가 3개국 시장 모두에서 상대 격차 수치가 양수(+)값을 보여 관심도 점유율이 낙관적인 것으로 나타났고, 글로벌 시장에서 보인 것처럼 실제 매출액 점유율은 시간이 지남에 따라 점점 증가하여 관심도 점유율 변화가 선행함을 잘 보여주었다.

<Table 8> Market Share in Germany

Year	Enbrel (Interest), %	Enbrel (Sales), %	Simple Difference	Relative Gap
2013	23.7	31.2	-7.50	-0.27
2014	22.1	29.7	-7.60	-0.29
2015	22.6	29.5	-6.90	-0.26
Average			-7.33	-0.28
Year	Remicade (Interest), %	Remicade (Sales), %	Simple Difference	Relative Gap
2013	25.1	19.9	5.20	0.23
2014	23.7	20.3	3.40	0.15
2015	23.6	18.9	4.70	0.22
Average			4.43	0.20
Year	Humira (Interest), %	Humira (Sales), %	Simple Difference	Relative Gap
2013	51.2	48.8	2.40	0.05
2014	54.2	49.9	4.30	0.08
2015	53.9	51.6	2.30	0.04
Average			3.00	0.06

<Table 9> Market Share in France

Year	Enbrel (Interest), %	Enbrel (Sales), %	Simple Difference	Relative Gap
2013	24.9	27.7	-2.80	-0.11
2014	24.6	26.3	-1.70	-0.07
2015	22.5	24.6	-2.10	-0.09
Average			-2.20	-0.09
Year	Remicade (Interest), %	Remicade (Sales), %	Simple Difference	Relative Gap
2013	34.9	34.1	0.80	0.02
2014	33.6	33.8	-0.20	-0.01
2015	33.1	32.4	0.70	0.02
Average			0.43	0.01
Year	Humira (Interest), %	Humira (Sales), %	Simple Difference	Relative Gap
2013	40.1	38.2	1.90	0.05
2014	41.8	39.9	1.90	0.05
2015	44.4	43.0	1.40	0.03
Average			1.73	0.04



### 3.3 바이오시밀러 제품의 시장 점유율 추정

바이오시밀러 제품의 경우는 후발 제품으로 오리지널 제품의 시장을 침투하는 품목이기 때문에 후발 제품의 관심도 데이터가 아직은 충분하지 않아 시장 점유율 분석이 통계적으로 유의하지 않다. 하지만 후발 제품이 향후 점유율을 얼마나 높여 시장에 성공적으로 침투할 것인지를 추정하는 것은 흥미로운 주제이다. 본 연구의 3.1절, 3.2절에서 TNF-α 저해제의 글로벌 상위 3개 제품의 매출액 점유율이 관심도 점유율과 높은 상관관계와 유사도를 갖는 것과 관심도 점유율은 매출액 점유율과는 다르게 후발 주자의 경우도 충분한 빅데이터 자료를 수집 분석할 수 있다는 사실을 사용하여 오리지널 제품과 후발주자인 바이오시밀러 제품의 매출액 점유율을 추정하고자 한다.

이를 위해 바이오 의약품 시장규모가 가장 큰 미국 시장을 선택하여, 오리지널 제약사의 글로벌 상위 3개의 TNF-α 저해제 중 가장 먼저 바이오시밀러 허가가 이루어진 레미케이드 제품의 바이오시밀러 제품을 연구 대상으로 선정하였다. 미국 시장에 출시된 레미케이드에 대한 바이오시밀러 제품은 셀트리온사의 인플렉트라와 삼성바이오에피스의 렌플렉시스가 있다. 엔브렐, 휴미라의 경우 국내 제약사 중 삼성바이오에피스의 바이오시밀러 제품만 최근 미국식품의약국으로부터 판매 허가를 받았으나, 너무 최근(각각 2019년 4월, 2019년 7월)이라 관심도 데이터가 충분히 누적되지 않아서 본 연구의 대상에 포함시키지 않았다.

<Table 10>은 미국식품의약국에서 제공하는 퍼플북(Purple book)[9]을 통한 레미케이드, 인플렉트라, 렌플렉시스 각각의 판매 허가일과 미국 시장에서의 제품 출시일을 보여준다. 이를 바탕으로 바이오시밀러 제품 출시 전·후 시점인 2015년 1월부터 2017년 12월까지 구글 트렌드를 사용하여 <Table 11>처럼 레미케이드 계열 3개 제품에 대한 관심도 점유율 분석을 분기단위로 수행하였다.

<Table 11>에 나타난 3년치 12분기 관심도 점유율 데이터를 사용하여 미국 시장에서의 오리지널 제품과 바이오시밀러 제품의 실제 매출액의 시장 점유율을 추정하였다. 추정하는 목표 시점은 레미케이드, 인플렉트라, 렌플렉시스 간에 본격적인 경쟁이 시작되며 3개 제품 모두의 매출액 집계 가능한 2018년도로 설정하였다. 즉, 2015년부터 2017년까지의 분기별 관심도 점유율 데이터를 사용하여 2018년도 1분기의 관심도 점유율을 예측하고, 이것을 2018년도 1년 간의 매출액 점유율로 추정하는 모형을 구축하였다. 관심도 예측 모형으로는 관심도 점유율 실측치 자료의 시간 성분과 불규칙 성분을 반영할 수 있는 단순 지수 평활 모형(Single Exponential Smoothing Model)을 사용하였다. 이는 현시점에 가까운 실적치에 큰 비중을 두면서 과거로 거슬러 올라갈수록 그 비중을

<Table 10> FDA Approval Date and Market Release Date of Remicade and Biosimilars

Product Name	Company	FDA Approval	US Market Release
Remicade	J&J	1998. 08	1998. 10
Inflectra	Celltrion	2016. 04	2016. 12
Renflexis	Samsung Bioepis	2017. 05	2017. 07

<Table 11> Interest based Market Share of Remicade and Biosimilars(2015~2017)

Year	Quarter	Remicade (Interest), %	Inflectra (Interest), %	Renflexis (Interest), %
2015	1Q	98.2	1.8	0.0
	2Q	97.4	2.6	0.0
	3Q	98.1	1.9	0.0
	4Q	97.2	2.7	0.2
2016	1Q	97.7	2.2	0.1
	2Q	94.2	5.5	0.3
	3Q	96.4	3.5	0.1
	4Q	93.1	6.6	0.3
2017	1Q	92.6	7.0	0.3
	2Q	91.3	6.9	1.8
	3Q	89.6	6.4	4.0
	4Q	90.5	6.9	2.6

적게 주는 지수 가중 이동 평균법으로 시계열 기반의 데이터 예측에 적합하다[15]. 구체적인 관심도 점유율 예측 및 매출액 점유율 추정 수식은 아래와 같다.

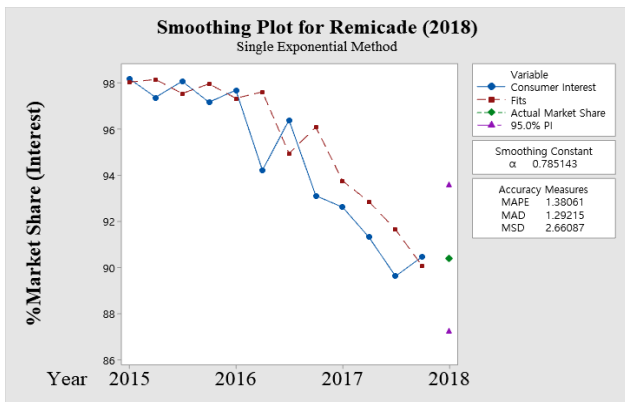
(단순 지수 평활 모형의 수식)

$$F_t = \alpha \times (C_{t-1}) + (1-\alpha) \times F_{t-1}$$

$\alpha$ : 지수평활계수  $0 \leq \alpha \leq 1$ ,  $C_{t-1}$ : 분기( $t-1$ )의 관심도 점유율 실측치,  $F_{t-1}$ : 분기( $t-1$ )의 관심도 점유율 예측치)

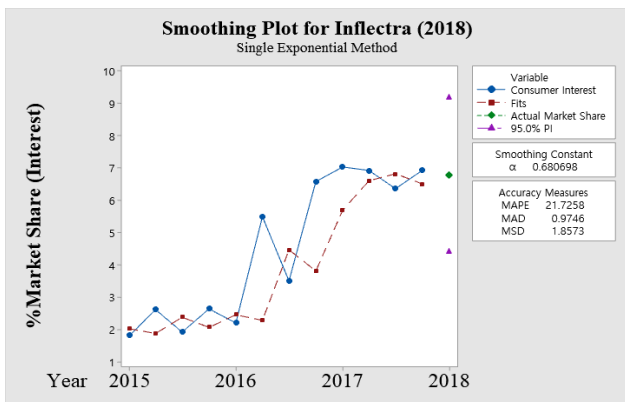
2015년부터 2017년까지 3년(12분기)의 관심도 점유율 자료를 사용하여 2018년 1분기의 관심도 점유율 예측치인 F13를 예측한 후, 최종적으로 이를 2018년 1년간의 매출액 점유율로 추정하였다. 단순 지수 평활 모형에서는  $\alpha$ 값이 클수록 예측치는 수요 변화에 빠르게 반응하며 예측의 감응도를 높인다. 여러 개의  $\alpha$ 값 중에서 예측 오차를 최소로 하는  $\alpha$ 값을 선택하는데, 모델 분석에는 미니탭 19 프로그램을 사용하였다. <Figure 8>, <Figure 9>, <Figure 10>은 <Table 11>을 기반으로 각각 레미케이드, 인플렉트라, 렌플렉시스의 분기별 관심도 점유율 자료를 예측 함수에 대입해 구한 관심도의 실측치와 예측치를 나타낸 것이다.





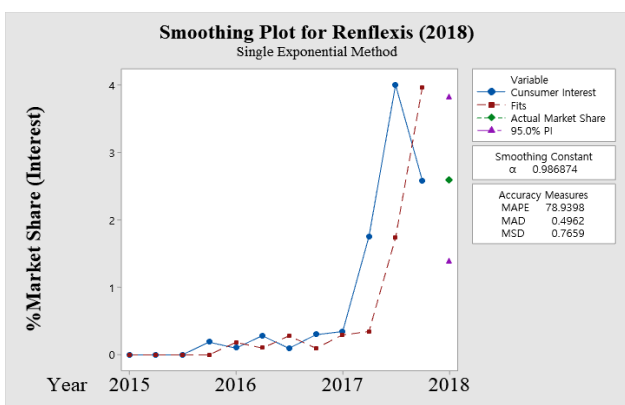
Period	Forecast	Lower	Upper
13	90.39	87.23	93.56

<Figure 8> Exponential Smoothing Model for Market Share Estimation (Remicade)



Period	Forecast	Lower	Upper
13	6.80	4.41	9.19

<Figure 9> Exponential Smoothing Model for Market Share Estimation (Inflectra)



Period	Forecast	Lower	Upper
13	2.60	1.38	3.81

<Figure 10> Exponential Smoothing Model for Market Share Estimation (Renflexis)

레미케이드, 인플렉트라, 렌플렉시스에 대하여 단순 지수 평활법을 통해 2018 년도 매출액 점유율을 추정하였을 때 각각 90.39%, 6.80%, 2.60% 값으로 추정되었으며, 각각의  $\alpha$  값은 0.79, 0.68, 0.99의 높은 감응도를 보여 예측치가 수요에 빠르게 반응하는 것을 보여주고 있다. 평균 절대편차(MAD, Mean Absolute Deviation)의 경우 레미케이드 (1.29%), 인플렉트라(0.97%), 렌플렉시스(0.50%)로 나타나, 매년 추정치가 실측치보다 약 1%p의 크지 않은 오차를 보여주었다.

구글 트렌드에 의한 관심도 점유율에 기반하여 매출액 점유율이 추정 함수대로 전개되는지를 검증하기 위해 수집할 수 있는 인플렉트라, 렌플렉시스의 실제 매출액 정보를 통하여 제안한 추정모델을 검증하였다. 오리지널 제품인 레미케이드의 경우 년 단위의 실제 매출액은 알려져 있으나, 바이오시밀러 제품의 경우 시장 진입 초기에는 기업들이 매출액 정보에 민감하여 공개하지 않거나 규모도 미미하여 공개된 자료가 제한적이었다. 인플렉트라의 경우 2018년 미국 시장 매출액이 259 M USD[3]으로 알려져 있고, 렌플렉시스의 경우 2018년 매출액이 94 M USD로 추정된 바 있다[12]. 이를 바탕으로 같은 해 3,664 M USD 매출액[3]을 보인 레미케이드를 포함하면 레미케이드, 인플렉트라, 렌플렉시스 각각이 91.21%, 6.45%, 2.34%의 실제 매출액 점유율을 가진 것을 확인하였다. 이와 같은 실제 매출액 점유율을 본 연구가 제안한 관심도 점유율 추정 함수를 통해 구한 2018년 매출액 점유율 추정치와 비교하면, 실제 매출액 점유율은 각각의 추정치의 95% 예측구간 (Prediction Interval) 범위 이내에 들었다. 또한 이들은 각각 관심도 점유율 기준으로 -0.82%p, 0.35%p, 0.26%p 수준의 차이를 보이고, 상대 격차 값이 -0.009, 0.053, 0.105로 매우 낮아 추정이 정확함을 보여주었다.

이상과 같이 미국 시장에 출시된 오리지널 제품인 레미케이드와 바이오시밀러 제품인 인플렉트라, 렌플렉시스의 매출액 점유율을 비교한 결과, 구글 트렌드에 나타나는 소비자 관심도 데이터가 실제 매출액 점유율 추정에 효과적으로 적용될 수 있음이 확인 되었다.

#### 4. 결론 및 시사점

본 연구를 통해 알아본 글로벌 상위 3개의 TNF- $\alpha$  저해제 제품(Enbrel, Remicade, Humira)과 바이오시밀러 제품 (Inflectra, Renflexis)에 대하여 수집 분석한 구글 트렌드에 의한 관심도 점유율은 년도별, 국가별로 매출액 점유율과 매우 강한 상관관계와 유사도를 보여 실제 시장 점유율 분석과 추정이 가능함을 보여주었다. 이는 의사나 환자 등 소비자의 관심도가 중요한 의약품 시장에서의 관심도 점유율 분석을 통한 빅데이터의 기업 경영 활용이 가능함을

보여준 결과이며, 독과점 형태의 시장지배 구조를 갖는 바이오의약품 치료제 시장에서 품목별 시장 점유율 분석이 시장 전략 수립에 효과적일 것이라는 함의를 준다.

본 연구에서 살펴본 오리지널 제품의 경쟁시 휴미라 사례처럼 후발 주자가 우수한 약효, 기술적 진보에 더해 공격적인 마케팅 결과로 소비자 관심도가 선행적으로 확대된다면 매출액 점유율의 확대도 나타난다는 것을 알 수 있었다. 또한 제안한 추정식에 의해 오리지널 제품인 레미케이드에 대해 후발주자인 인플렉트라, 렌플렉시스 등 바이오시밀러 제품은 FDA 승인 1, 2년 만인 2018년에 각각 6.80%, 2.60%의 매출액 점유율을 달성한 것으로 추정이 되었고 실제로는 6.45%, 2.34%의 매출액 점유율을 보여 95% 예측구간 이내의 추정이 가능하였다. 따라서 바이오시밀러를 출시하는 후발주자들이 오리지널 제약사 제품의 기존 시장을 잠식함에 있어서, 구글 트렌드의 관심도 점유율 추정과 이를 바탕으로 한 관심도 고양 전략은 매우 효과적이라고 판단한다.

구글 트렌드와 같은 빅데이터 분석도구는 전세계를 대상으로 하여 소비자의 관심도를 시기적, 지역적, 제품별로 추적이 가능하다는 속성이 장점이다. 따라서 빅데이터 분석도구가 관심도에 민감한 제품의 매출액 점유율 분석과 추정에 활발히 사용된다면 학문적으로나 기업 경영에 있어서 의미있는 시도가 될 것으로 판단한다.

본 연구에서는 구글 트렌드를 통하여 소비자 관심도 데이터를 월별 또는 분기별로 수집하여 경쟁제품들의 관심도 점유율을 분석할 수 있었다. 하지만 매출액 점유율 분석시에는 의약품의 특성상 년간으로 누적된 매출액 데이터를 사용한 분석은 가능하나 더 이상 세분화된 매출액 자료에 대한 접근은 제한적이므로 분기별 또는 월별 분석은 불가능하였다. 향후 매출액 데이터를 분기별, 월별로 추적할 수 있는 제품에 대해 연구한다면 좀 더 빠르게 시장 점유율 변화를 분석할 수 있을 것이다.

또한 영문 제품명만을 키워드로 하는 것이 아니라 부작용 등을 포함한 검색 키워드 최적화와 브랜드에 대한 감성 분석(sentiment analysis) 등과 결합한 연구는 빅데이터 분석을 통해 소비자들의 반응을 기업 경영에 활용하는 연구의 발전에 기여할 것이다.

## Acknowledgement

This work was supported by the National Research Foundation Korea(NRF) grant funded by the Korean government (MSIT)(No.2018R1D1A1B07050139).

## References

[1] Bae, E.Y., Determinants of drug market share, *The*

*Korean Journal of Health Economics and Policy*, 2000, Vol. 6, No. 2, pp. 1-30.

[2] BIOSPECTATOR, [http://www.biospectator.com/view/news\\_view.php](http://www.biospectator.com/view/news_view.php).

[3] BIOSPECTATOR, [http://www.biospectator.com/view/news\\_view.php?varAtcId=7055](http://www.biospectator.com/view/news_view.php?varAtcId=7055).

[4] Biotech Policy Research Center, Strategic insights into biopharmaceuticals industry, Frost & Sullivan Analysis, 2019. 3.

[5] BP Technology Trade, Industrial analysis report of biologics, 2018, pp. 2-29.

[6] Chamberlin, G., Googling the present, *The Labour gazette*, 2010, Vol. 4, No. 12, pp. 59-95.

[7] Choi, H. and Varian, H., Predicting the present with Google Trends, *Economic Record*, 2012, Vol. 88, No. 1, pp. 2-9.

[8] Dugas, A.F., Hsieh, Y.H., Levin, S.R., Pines, J.M., Mareiniss, D.P., Mohareb, A., and Rothman, R.E., Google flu trends : Correlation with emergency department influenza rates and crowding metrics, *Clinical Infectious Diseases*, 2012, Vol. 54, No. 4, pp. 464-469.

[9] FDA, <https://www.fda.gov/drugs/therapeutic-biologics-applications-bla/purple-book-lists-licensed-biological-products-reference-product-exclusivity-and-biosimilarity-or>.

[10] Fortune Korea, <http://www.fortunekorea.co.kr/news/articleView.html?idxno=11536>.

[11] Korea biologics and biohealth information assistance system, <http://www.kobics.or.kr/UBFCtr.do>.

[12] Lee, B. and Park, W.Y., Company update : Samsung biologics, Samsung Securities, 2017, p. 6.

[13] Lee, J.B. and Paek, D.H., The effect of smartphone purchasing determinants on repurchase intention, *Journal of Society of Korea Industrial and Systems Engineering*, 2017, Vol. 40, No. 2, pp. 1-12.

[14] Magazzini, L., Pammolli, F., and Riccaboni, M., Dynamic competition in pharmaceuticals Patent expiry, generic penetration, and industry structure, *European Journal of Health Economics*, 2004, Vol. 5, pp. 175-182.

[15] Naver, <https://blog.naver.com/sigmagil/221502514892>.

[16] Oh, K.Y., Kim, B.H., and Kim, T.G., Estimating market share function of generic drugs after patent expiration in Korea, *The Korean Journal of Health Economics and Policy*, 2007, Vol. 13, No. 1, pp. 117-138.

[17] Qin, S. J., Process data analytics in the era of big data, *AIChE Journal*, 2014, Vol. 60, No. 9, pp. 3092-3100.

[18] Shin, M.S., Park, M.G., and Bae, S.H., Nano technology

- trend analysis using google trend and data mining method for nano-informatics, *Journal of Society of Korea Industrial and Systems Engineering*, 2017, Vol. 40, No. 4, pp. 237-245.
- [19] Sohn, J.W. and Kim, J.K., Attributes of social networking services : A classification and comparison, *Journal of Society of Korea Industrial and Systems Engineering*, 2017, Vol. 41, No. 1, pp. 24-38.
- [20] Stress and Coping Strategies among Commuter and Non-commuter Students, [dissertation], [Kedah, Malaysia] : Universiti Utara Malaysia, 2013.
- [21] Youn, S. and Cho, H.C., Nowcast of TV market using google trend data, *Journal of Electrical Engineering and Technology*, 2016, Vol. 11, No. 1, pp. 227-233.
- [22] Yu, J.P. and Lee, B.U., Forecasting company sales and stock price using google trend : Focusing on the keywords of BMW and Mercedes-Benz, the convergent research society among humanities, *Sociology, Science, and Technology*, 2018, Vol. 8, No. 10, pp. 491-501.

**ORCID**Ki Tae Bong | <http://orcid.org/0000-0003-1729-5397>Heesang Lee | <http://orcid.org/0000-0002-1415-8410>