

치위생 분야의 조사연구 논문 통계분석과 결과 해석

김혜영 · 조영식¹ · 황수정^{2†}

고려대학교 치기공학과, ¹남서울대학교 치위생학과, ²건양대학교 치위생학과

Statistical Analysis and Interpretation of Survey Results in Dental Hygiene

Hae-Young Kim, Young-Sik Cho¹ and Soo-jeong Hwang^{2†}

Department of Dental Laboratory Science and Engineering, College of Health Science, Korea University, Seoul 136-856, Korea

¹Department of Dental Hygiene, Namseoul University, Cheonan 331-707, Korea

²Department of Dental Hygiene, College of Medical Science, Konyang University, Daejeon 302-718, Korea

Abstract This study aimed to propose appropriate methods of statistical analysis and interpretation of survey results in dental hygiene. By using examples of published articles in the Journal of Dental Hygiene Science from 2003 to 2008, we made some suggestions to improve scientific and clear writing of the contents; purpose of the study, subjects and methods, statistical analysis, tables, and interpretations.

Key words Data analysis, Dental hygiene, Statistics

서 론

한국치위생과학회지는 우리 나라 치위생 분야에서 최초로 유일하게 학술연구재단 등재학술지에 선정되었으며 치위생학의 학문적 발전의 기반을 다지는 데 크게 공헌하고 있는 치위생학계의 대표적인 학술지 중 하나이다. 이 논문은 2003년에서 2008년 사이에 발간된 한국치위생과학회지에 수록된 전체 논문의 약 70% 정도를 차지하는 조사연구 논문의 연구목적, 연구대상 및 연구방법, 통계분석 방법 및 결과해석 등을 비판적으로 검토하여 보다 과학적이고 명료한 기술방법과 분석 및 해석방법을 제시함으로써 한국 치위생 분야의 조사연구 논문의 질적 발전에 기여하고자 하였다. 검토된 논문의 일부를 변형하여 본문에 예시된 사례로 활용하였고 편의상 개별 사례의 구체적인 출처는 밝히지 않았다.

본 론

1. 연구 목적 관련 검토 사항

한국치위생과학회지에 게재된 조사연구 논문들은 다양한 주제를 다루고 있어 구강보건 의식과 실태 및 행태, 구

강건강관련 삶의 질, 구취, 치과진료 시 공포감, 스트레스, 치과위생사의 이미지나 직무 관련 사항, 환자의 진료만족도, 의료사고 등이 포함되었다. 연구의 내용 상 기술적인 주제를 다룬 논문이 한국치위생과학회지에 수록된 논문의 대부분을 차지하여 “000에 대한 인식도 조사연구”, “000의 인식 및 실천도” 등의 제목으로 표현된 논문들을 예시할 수 있다. 상대적으로 분석적인 주제를 담은 조사연구 논문은 적었고 “000의 영향 요인”, “00와 00의 관련성 연구” 등의 제목으로 표현 되고 있었다. 기술적인 내용을 담은 논문에서는 가설을 설정하지 않는 경우가 많아 많은 논문에서 연구가설이 설정되지 않고 “00에 관하여 알아보고자” 등의 연구목적으로 표현되었다. 분석적인 주제를 다루는 조사연구논문에서 연구가설은 연구문제에 대한 연구자 자신의 잠정적인 결론 또는 해답을 서술한 것으로 일반적으로 측정 가능한 둘 이상의 변수들 간의 관계를 예측한 서술로, 명확하고 구체적이어야 하고 검증이 가능한 형태로 표현되어야 한다는 대원칙을 준수하여야 한다¹⁾.

2. 조사 대상 관련 검토 사항

1) 모집단과 표본에 대한 기술

조사대상에 관한 기술에서 표본 선정은 연구대상 모집단과 표본에 대한 개념을 포함하여 서술하는 것이 필요하다. 연구대상 모집단(population)이란 연구결과에 대하여

[†]Corresponding author

Tel: 042-600-6381

Fax: 042-600-6382

E-mail: denthwang@konyang.ac.kr

입증하고자 관심을 갖는 전체 집단으로 연구의 결론을 내리고자 하는 집단을 말하며 표본(sample)이란 모집단의 일부로서 연구대상 모집단에 관한 정보를 얻고자 직접적으로 조사를 하는 집단을 말한다^{2,3)}. 따라서 연구조사의 직접적 대상이 되는 표본은 연구대상 모집단을 대표하는 표본으로서 구체적으로 서술하여야 한다. 한국치위생과학회지에 나타난 표본 관련 설명을 살펴보면 “전국 치과 병의원 종사자 95명” 또는 “수도권 소재 치과에 근무하는 치과위생사를 대상으로” 등의 표현이 대부분으로 차지하여 연구대상 모집단이 무엇이며 어떤 방법으로 표본을 선정하여 연구대상 모집단을 대표하는 표본을 추출하여 조사하게 되었는지 등의 내용이 구체적으로 서술되어 있지 않는 경우가 많았다. 관심을 가지고 결론을 내리고자 하는 연구대상 모집단을 규정하고 단순무작위 추출 등 대표성 있는 표본을 뽑아내는 원칙적인 방법을 사용하여 연구대상의 설명으로 제시하는 것이 요청된다.

표본 선정에 관한 서술 상 조사 대상에 제한 구체적인 묘사가 필요한 경우가 다소간 발견되었는데 예를 들면 “00시 소재 고교생을 대상으로” 라는 기술에 대하여 00시의 모든 고교가 연구의 최종 해석을 하고자 하는 대상이 모집단이 되는 것인지 아닌지가 분명하지 않다는 문제점을 지적할 수 있고 00시의 몇 개 고교 중에서 어떤 학년을 대상으로 어떤 표본 추출 방법을 이용하여 추출하였다는 것을 밝혀 주어야 독자가 이 표본이 연구대상 모집단을 대표하는 대표성이 있는 표본인지를 판단하여 분석결과와 최종해석을 어떤 단위로 하여야 할 것인가를 선택할 수 있다. “시 및 군 지역의 성인 500명” 이란 표현도 마찬가지로 편의적 선택 방법으로 한 것으로 간주될 수 있고 시 및 군지역의 성인 인구를 대표하는 대표성을 가지는 표본인지, 결과 해석을 시 및 군지역의 성인을 대상으로 적용할 수 있는 지에 대하여 의문을 갖게 하는 묘사이다.

조사연구에 의한 증거의 신뢰성에 대한 평가는 그 연구가 어떠한 방법으로 선정된 표본인가에 대하여 크게 의존한다. 질적 연구방법을 채택한 경우와는 달리 양적 연구방법을 채택한 경우, 개별적인 일례 자체로는 큰 증거가 될 수 없을 것이다. 편의 표본 추출에 의거한 조사결과는 (대표성이 없는) 대상 집단의 일부에 국한된 결과로 의심을 받을 수 있다. 대표성 있는 표본에 의거한 조사, 표본의 크기가 크고 여러 번 반복하여 동일한 결과가 나오는 경우 등이 연구가설에 대한 강력한 증거로 채택될 수 있다.

2) 무응답과 응답률의 제시

선정된 표본 중 조사과정에서 발생하는 무응답은 중요한 오류의 원인으로 지적되며 특히 높은 무응답률은 비관찰오류(nonobservational error) 중 가장 심각한 문제로 지적되고 있다. 선정된 표본의 수와 조사 결과의 응답수, 무

응답수 및 응답률 등은 반드시 보고하여야 하는 항목들이다. 낮은 응답률 뿐 아니라 응답률이 높은 경우에도 무응답은 오류의 원인이 될 수 있으므로^{3,4)} 무응답자에 대한 분석을 하여, 조사된 표본이 조사되지 않은 부분과 다르지 않고 무응답의 이유가 연구주제와 관련이 있지 않다는 것을 입증해 주는 것이 필요하다. 한국치위생과학회지에 수록된 논문에 나타난 연구대상 관련 바람직하지 못한 서술을 예로 들면 “회수된 설문지 200부 중 자료처리에 부적합한 40부를 제외한 총 160부를 자료분석에 이용하였다”와 같이 최초에 선정된 표본수를 밝히지 않고 최종 설문문에 응한 응답자의 수만 제시하는 것으로 위의 서술에서는 무응답의 규모나 응답률이 나타나 있지 않고 응답한 표본이 선정된 표본과 동질적이라는 근거가 제시되어 있지 않다. 좀 더 바람직한 예로는 “선정된 1,500개 표본 치과에 설문을 실시하여 그 중 1,000 개의 응답을 얻었고 (응답률 66.7%) 그 중 주요한 관심항목에 결측치가 있는 30개 설문을 제외한 총 970개의 응답을 분석에 사용하였다(최종 64.7%)”와 같이 최초 선정 표본 수, 응답 표본수 및 응답률, 및 자료 분석에 사용된 표본 수 등을 모두 제시한 예를 들 수 있다.

3) 한 두 개의 표본집단 사용 시 고려사항

편의상 조사 대상 표본을 한 두 개의 집단에서 선정하는 경우가 있다. 예를 들면 한 두 개의 치과의원에 내원한 환자 집단에 대한 조사로서 “00년 한 치과의원에 내원하여 영구치를 발거한 1000명의 환자를 대상으로”라고 표현된 경우가 이에 해당한다. 이러한 경우에도 궁극적으로는 전체 치과의원이 최종적인 관심의 대상이 될 것이므로, 연구대상 모집단과 표본으로 조사된 특정 치과의원의 특징을 비교하거나 표본 집단의 배경 등을 구체적으로 서술하여 전체 관심 대상 집단 속에서의 위치를 이해하고 후속연구에서 참조 가능하게 하여야 할 것이다. 예를 들면 조사된 치과의원의 규모, 소재하는 지역의 특성(도시 또는 농촌, 상업지역 또는 주거지역 등), 및 내원하는 환자의 특성(가족 또는 회사원 중심 등)을 기술한다면 향후 같거나 다른 특성을 가지는 표본에 대하여 이루어지는 연구에서 참고할 수 있고 연구결과를 비교, 평가할 수 있는 자료로 활용 가능할 것이다.

3. 조사 방법 관련 검토 사항

1) 이론적 개념들에 의거한 조사항목 선정

설문조사의 내용을 선정할 때의 일반적인 원칙은 연구조사하려는 내용 전체에 대한 이론적 개념틀(conceptual framework)을 먼저 세운 다음 그 이론적 개념틀에 의거한 조사항목을 선정하고 가능하다면 기존 연구에서 전체 그 의미나 신뢰도가 검증된 조사항목을 선정하여 참조 문헌을 제시하는 것이다. 따라서 각 항목이 가지는 전체 개념틀에서의 의미가 명확하게 제시되고 이에 따라 각 항목의

조사결과가 해석되어야 한다. 그러나 한국치위생과학회지에 수록된 많은 조사연구 논문에서는 사전에 이론적 개념들을 제시하지 않거나 기존 연구에서 검증된 조사항목을 사용하고 출처와 의미를 정확히 밝히지 않는 경향이 발견되었다.

2) 조사항목의 서술

설문조사 항목에 대한 묘사는 측정설문에 대한 구체적 서술로서 무엇을 어떻게 물었고 결과를 어떻게 해석하여야 하는지 알려주어야 한다. 한국치위생과학회지에 수록된 논문 중에서 발견된 예를 들면, “구강건강신념 지표는 리커트 척도로 측정하였다”로만 기술하고 구체적으로 어떤 점수가 어떤 의미로 쓰였다는 것을 제시하지 않은 채 연구결과에서 평균값으로 요약하여 제공한 예가 있었다. 이러한 경우 독자는 구강건강신념이 구체적으로 무엇을 의미하는지를 알 수 없을 뿐 아니라, 결과로 요약된 숫자도 의미를 알 수 없게 된다. 보다 바람직한 예를 제시하면, “구강건강인식도는 자신의 구강건강이 어떠하다고 생각하는 지를 질문하여 (1)매우 좋음, (2) 좋음, (3) 그저그러함, (4) 나쁨, (5) 매우 나쁨 등의 답지에서 고르도록 하는 방법으로 조사함”과 같이 문항의 의미와 각각의 값의 의미를 구체적으로 서술하여야 독자가 그 의미를 파악할 수 있을 것이다. 연구방법에서 답지의 내용을 직접 서술하지 않고 결과표에 답지의 내용을 표기하여 독자가 알아볼 수 있도록 하는 방법도 가능하다.

3) 통합지표의 활용 시 고려사항

여러 문항을 통합하여 어떤 의미를 부여한 것을 통합지표라고 하며, 구강건강 관련 연구에서 자주 활용되는 통합지표의 예로 주관적 구강건강관련 삶의 질을 평가하기 위한 구강건강영향지수(Oral Health Impact Profile)를 들 수 있다. 한국치위생과학회지에 수록된 논문 중 일부의 경우, 이러한 통합지표를 사용할 때 지표를 구성하는 각 항목에 대한 설명이나 통합지표의 생성 과정 등에 대한 설명이 연구방법에서 생략된 채 연구결과에서 평균치 등을 제시하여 그 통합지표의 의미를 이해할 수 없는 경우가 발견되었다. 통합지표 사용 시에는 먼저 통합지표를 구성하는 각 문항에 대한 설명이 선행되어야 하며 또한 문항의 값이 의미하는 것이 동일한 방향으로 정리된 후 Cronbach's alpha 등을 이용한 신뢰도 검사를 거쳐 통합지표의 생성 근거를 제시 하고 통합지표의 생성 과정과 해석 방법을 제시하는 것이 일반적으로 요구 된다. 통합지표의 생성과정과 해석 방법에 대한 구체적인 서술의 예로 “OHIP-49항목의 값을 합하여 OHIP score로 한다. OHIP score는 0~196 사이의 값을 가지며 0은 구강건강과 관련된 문제로 (좋지 못한)영향을 전혀 받지 않는다는 경우이며 196은 모든 열거된 구강건강과 관련된 문제를 매우 자주 경험하였다는 의미로 해석한다.”와 같이 설명한

것을 예시할 수 있다. 또한 통합지표의 신뢰도가 입증된 기존 연구결과가 있다면 이를 적극 활용하고 참고문헌으로 제시하여야 할 것이다.

4) 설문조사와 구강검사 방법의 기술

한국치위생과학회지에 수록된 조사연구 논문을 살펴본 결과, 많은 사례에 있어서 설문조사 방법의 기술에 있어서 보다 구체적인 기술이 필요한 것으로 사료되었다. 응답자에 의한 자기기입법으로 시행한 것인지 또는 면접자에 의한 면접법에 의하여 시행한 것인지 구체적인 조사방법의 묘사가 필요한 경우가 다수 발견되었다. 또한 면접법에 의하여 수행된 연구라면 훈련된 면접자가 동일한 기준으로 설문하였다는 것을 10%의 이중응답을 수행하여 면접자간 또는 표준면접자와 면접자의 일치도를 검증하는 과정이 필요하다⁵⁾.

구강검사방법의 기술에 있어서는 검사 기준의 명확한 제시, 검진자 훈련 과정의 보고와 그 결과로 얻은 검진자 내 및 검진자간의 신뢰도 보고 부분이 제시되어 있지 않은 경우가 많았다. 따라서 실제 검진 대상과 유사한 20명 가량을 대상으로 하여 반복검진하여 얻은 카파계수나 급내상관관계(intraclass correlation)의 값을 보고하여야 한다. 더불어 현장에서는 10%의 대상자를 이중검진하여 치아우식 관련 지표, 치주질환 관련 지표 및 현존치아수 등 구강건강상태에 대하여 검진자간 신뢰도 또는 표준검진자와 검진자간의 신뢰도를 보고하여야 한다.

4. 통계분석 관련 검토 사항

1) 자료의 분포와 분석방법의 선행조건 고려

한국치위생과학회지에 나타난 논문들을 살펴보면 대부분의 연구자들이 분석대상 자료의 분포를 검토하지 않은 상태에서 분석방법의 기본 선행조건의 충족 여부를 탐구하지 않고 곧바로 자료 분석방법을 적용하는 경향을 보임을 지적할 수 있다. 이상치, 정상적 분포에서 벗어난 데이터가 있는지를 먼저 검토하여야 하며 연속변수의 경우 모수적 분석방법을 사용하려면 대상 자료의 모집단이 정규분포 한다는 가정에 적합한 지 여부를 선행하여 검증하고 그 결과에 따라 모수적 방법 또는 비모수적 방법을 선택하여야 한다. 예를 들어 t-test를 적용하려면 모집단의 정규성 검정 결과에서 정규분포를 가정할 수 있어야 하고 또한 모분산의 동질성 검정을 한 후, 그 결과에 따라 공통분산을 사용하는 방법을 사용하거나 사용하지 않는 방법 (Welch's test)을 선택하여야 한다. 일원분산분석의 경우에는 모집단의 정규분포의 가정에 만족하고 모분산의 동질성이 인정되어야 일반적으로 적용할 수 있으며, 만약 그 조건에 만족하지 못하면 Tamhane's test 등 특수한 사후분석방법을 선택하거나 비모수적 방법을 적용하여야 한다⁶⁾.

범주형 변수의 연관성 분석 시 많은 경우 카이제곱검정

을 이용하는데 카이제곱검정의 결과를 신뢰할 수 있는 기본조건은 기대치 5이하인 셀이 20%이하이어야 하며 빈도가 0인 셀이 없어야 한다는 것인데, 이러한 조건을 지키지 않은 상태에서 분석하여 카이제곱검정 결과를 보고하는 경우가 다수 발견되었다. 만약 이러한 조건에 부합되지 않는다면 Fisher's exact test를 적용하여야 한다.

2) 이론적 토대에 의거한 연관성 분석

일반적으로 데이터의 연관성 분석은 사전 연구에 의하여 비교하고자 하는 두 개념의 연관성이 있다고 보고되었거나 또는 여타의 연관성이 있을 것이라는 합리적인 근거를 먼저 제시한 다음 시작하여야 한다. 한국치위생과학회지에 수록된 조사연구 논문의 많은 사례에서 보여지는 바와 같이 이론적 배경에 대한 사전 서술이 없는 상태에서 이변량 분석을 하고 유의한 연관관계를 보고하는 방식은 사전에 계획되어 분석을 하는 것이 아니라 일단 분석을 해 보고 결과를 찾아보려는 것으로 인식될 우려가 있다. 또한 분석에 의하여 통계적인 유의성이 발견된 경우라도 합리적인 이론적 토대 위에서 연관성을 설명하는 면이 부족한 경우가 많았다.

3) 기타 고려점

-일원분산분석의 결과가 유의하여 다수 집단의 평균값이 모두 동일한 것은 아니라는 결론이 났을 경우에 사후분석을 진행하여 구체적으로 어떤 집단간의 평균값이 서로 유의하게 다르고 어떤 집단간의 평균값이 다르지 않다는 결론을 내려야만 실제적으로 그 결과를 활용할 수 있다. 사후분석이 아예 진행되지 않았거나 사후분석의 결과가 적절하게 표현되지 못한 경우가 상당수의 논문에서 발견되었다.

-범주형 변수의 분석에서 층화된 범주형 변수의 연관성을 보고자 하는 경우, 예를 들면 남녀별로 구분된 자료에서 남녀 각각의 범주를 포함하여 치료방법과 반응결과의 관련성을 보고자 할 때 일반적 카이제곱검정 대신에 Mantel Haenszel 통계량을 이용하여 검정하여야 한다. 두 범주를 가지는 변수 간의 관련성을 요약할 때 오즈비를 이용하여 제시하는 것도 고려해 볼 만 하다⁷⁾.

-순서형 변수의 통계처리 시 5점 척도의 리커트 척도로 측정된 경우 리커트 척도를 연속변수로 취급하여 평균과 표준편차를 이용하여 비교하는 방법에 대해서는 가능하다는 의견과 바람직하지 못한 처리방법으로 사용하지 않는 것이 좋다는 의견이 공존하는 것으로 사료된다. 원칙적으로는 범주형 순서형 변수로 취급하여 통계분석법을 적용하는 것이 바람직한 것으로 사료된다.

5. 분석표 관련 검토사항

1) 결측치나 무응답 자료의 처리

한국치위생과학회지에 수록된 논문 중 결과표에 무응답

을 표시하고 무응답수를 포함한 백분율을 표기하여 무응답을 제외한 특정 범주의 분율을 알 수 없는 경우가 다수 발견되었다. 일반적으로 무응답이 특별한 의미가 있는 경우, 예를 들면 사회통념 상 답변하기 곤란한 질문에 대한 무응답 등 특별히 나타내어 주어야 하는 경우 이외에는 무응답은 제외하고 표를 작성하는 것이 좋다. 범주별 백분율은 무응답을 제외하고 100% 기준으로 계산하는 것이 바람직하며 그 결과 각 항목마다 대상수가 달라질 수 있다. 또한 반올림 과정 등으로 인하여 퍼센트 합이 100%가 되지 않는 경우가 발생할 수 있는데 이 경우에도 가능한 한 조정하여 100%로 맞추어 주는 것이 바람직한 것으로 사료된다.

2) 분석표 작성 시의 유의사항

표는 독립적으로 이해되고 의미를 가지도록 작성되어야 한다. 한국치위생과학회지에 나타난 일부 표 중에서는 독립적으로 이해하기 어려운 사례들이 발견되었다. 표의 제목은 담겨진 내용을 정확하게 나타내도록 작성하여야 하므로, 관심변수, 비교 집단, 관심을 가지는 변수 간의 관계 등이 정확하게 표현되도록 작성한다. 예를 들면 “t-검정결과”라는 제목이 관심변수와 관계에 대한 정보를 거의 포함하지 못하는 반면, “치료 전과 후의 발치군과 비발치 교정군의 치열 폭경의 평균 변화량의 비교”라는 제목은 관심 변수, 비교 집단 및 그들 간의 관계를 포함하는 정보를 잘 제공하고 있음을 알 수 있다.

한국치위생과학회지에 나타난 논문 중 일원분산분석을 이용하여 세 집단 이상의 평균비교를 실시한 상당수의 논문에서 사후검정 결과를 적절하게 표에 표현하지 못한 경우가 발견되었다. 또한 표에 사용된 약어와 기호에 대한 설명이 반드시 적절하게 이루어져야 한다.

6. 결과의 서술 관련 검토사항

몇몇 논문의 경우, 설문조사 결과의 서술 시 표현방법에 주의하여야 할 것으로 사료되었다. 예를 들면 설문조사 결과 치아우식증을 경험하였다고 응답한 사례에 대하여 “충치를 경험하였다고 응답한 남학생이 많았다”라고 표현하지 않고, “충치를 경험한 남학생이 많았다”으로 표현하면 구강검사를 통하여 조사된 결과로 잘못 이해될 수 있다. 상당수의 논문에서 고찰부분에서 이전에 설명한 결과부분을 반복할 뿐, 다른 자료와 비교하여 평가하거나 결과에서 나타난 특징을 심층적으로 탐구하여 진전된 결론에 이르는 등의 내용이 나타나지 않는 경향을 발견할 수 있었다.

요약 및 결론

한국치위생과학회지에 수록된 논문에 나타난 조사연구 논문을 연구방법과 통계분석 및 결과해설의 관점에서 비

관적으로 검토한 결과 다음과 같은 제언을 할 수 있다.

1. 기술적인 내용에 그치기보다는 분석적인 주제를 더 많이 연구하는 것이 바람직하다.
2. 모집단을 명확하게 제시하고 표본이 모집단을 대표하고 있음을 기술하여야 한다.
3. 이론적 근거에 기반한 설문 조사 항목의 선정과 구체적인 묘사가 요망된다.
4. 자료의 분포와 선행 조건에 기반 하여 통계분석방법을 선정하여야 하고 이론적 토대에 근거한 연관성 분석이 이루어져야 한다.

참고문헌

1. Lee JS: Writing Scientific Articles. Education Science, Seoul, pp.31-35, 2006.
2. Lee SW: Understanding Statistics. Free Academy, Seoul, pp.3, 2001.
3. Schwaffer RL, Mendenhall W, Ott RL. Elementary Survey Sampling, fifth edition. Duxbury Press, Belmont, pp.42-55, 1996.
4. Lohr SL. Sampling: Design and analysis, Duxbury Press, Pacific Grove, pp.2-6, 1999.
5. Kim JB et al.: Public Oral Health. Komoon Sa, Seoul, pp.399-415, 2008.
6. Bland M. An introduction to medical statistics, third edition. Oxford University Press, Oxford, pp.157-181, 2000.
7. Stokes ME, Davis CS, Koch GG. Categorical data analysis using the SAS®System. 2ndedition. Cary, NC, USA: SAS Institute Inc, pp.123-136, 2001.

(Received February 14, 2011; Revised February 14, 2011;
Accepted February 14, 2011)

