epidemiology biostatistics and public health
formerly ITALIAN JOURNAL OF PUBLIC HEALTH

Volume 18
Issue 1
September 2023

Milano University Press

epidemiology biostatistics and public health formerly ITALIAN JOURNAL OF PUBLIC HEALTH

Volume 18, Issue 1 | September 2023

CONTENTS

EDITORIAL

ORIGINAL ARTICLES

STATISTICAL METHODS

Milano University Press

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

EBPH is Back for a Global Audience

We are proud to announce the re-starting of the Journal Epidemiology Biostatistics and Public Health (EBPH), made possible thanks to the Milano University Press (MUP), the new publisher of the journal. MUP has included EBPH in the Diamond Journal collection and made available online all previous issues, too.

Volume 18, Issue 1 (2023) is ready as online publication, submissions are now open for Issue 2.

EBPH started through the merging of two previous journals, the Italian Journal of Public Health (founded by Walter Ricciardi) and the Journal of Biostatistics and Clinical Epidemiology (founded by Gianni Corrao), bridging together the areas of public health, epidemiology, and biostatistics. This was possible thanks to an initiative of the Italian Society of Medical Statistics and Clinical Epidemiology (Società Italiana di Statistica Medica ed Epidemiologia Clinica – SISMEC), together with eminent research groups in the area of Public Health. These have worked synergistically since 2011 to keep this important tool for disseminating research in the field of public health and methods to support it, namely epidemiology and biostatistics.

The disciplines integrate and guarantee that the journal develops as a further prestigious and reliable tool for healthcare professionals at any level: from those involved in clinical research to those entrusted to manage healthcare services. The recent experiences related to the Covid-19 pandemic and the consequent health emergency have also made clear the need for a rapid sharing of methods and findings. This will allow a larger international debate on relevant topics in epidemiology, biostatistics and public health.

EBPH, as the official journal of SISMEC, is aimed at disseminating research by working groups and initiatives from the Society and it is also an ideal instrument for the publication of articles in the public health area characterized by relevance of the research topics and innovation of the proposals.

The journal is open to publications coming from the global scientific community dealing with these issues.

President of SISMEC,
Rosaria Gesuita, PhD
Università Politecnica delle Marche

Past-president of SISMEC,
Paolo Trerotoli, MD
Università degli Studi di Bari Aldo Moro

Editors in Chief
Chiara de Waure, MD, MSc, PhD
Università degli Studi di Perugia

Carlo La Vecchia, MD
Università degli Studi di Milano

Assistant Editor
Rossella Bonzi, PhD
Università degli Studi di Milano

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

What is wrong with Chinese COVID-19 statistics?

Igor Nesteruk[(1) (2)]

(1) Institute of Hydromechanics, National Academy of Sciences of Ukraine, Kyiv, Ukraine.
(2) Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine.

CORRESPONDING AUTHOR: Igor Nesteruk, nesteruk@yahoo.com, ORCID 0000-0001-7250-2729.

SUMMARY

The media is reporting tens of millions of new daily Covid-19 cases in China in the final days of 2022. However, official statistics have recorded in mainland China only1.9 million cases since the start of the pandemic and stopped providing data after December 21, 2022. Results of SIR simulations showed that daily numbers of new cases stated to decline in December 2022. The contradictions in statistics and estimations are discussed. Millions of new daily cases in China look very unlikely.

Keywords: COVID-19 pandemic; epidemic waves; epidemic dynamics in China; mathematical modeling of infection diseases; the generalized SIR model; parameter identification; statistical methods.

According to the information from an internal meeting of China's National Health Commission (NHC) held on December 21, 2022, the 248 million people were infected with Covid-19 in the first 20 days of December with nearly 37 million new cases on a single day [1]. These figures contradict with official statistics, reflected in COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), [2]. According to the version of the JHU file, available on January 4, 2023, as of December 21, 2022 the accumulated numbers of cases were:1,909,905 in mainland China; 2,402,238 in Hong Kong; 8,624,680 in Taiwan;1,569 in Macao. Taking the sum of this figures, we have the value 12,938,392 that is 19 times lower than NHC estimation of the number of cases accumulated in the first 20 days of December 2022 and almost 3 times lower than its daily maximum figure.These huge discrepancies may indicate either completely incorrect NHC statistical data (which stopped be ingupdated after December 21, 2022, see [2] and Table1), or a new estimate of the real number of cases (the details of which we do not know).

The difference between the real and detected numbers of COVID-19 cases can be huge, [3,4], especially in countries with low testing level [4]. In mainland China the number of accumulated tests per capita TC was 6.46 already on April 11, 2022, [2]. The testing levels in Hong Kong are similar (TC=6.59 as of May 24, 2022, [2]), and approach to the highest values in other countries [2]. Thus, there is no reason to think that many cases in China was not detected even after November 30, when the Zero-Covid [5] restrictions began to loosen [6].

The epidemic dynamics in mainland China in December 2022 can be compared with the results of application of the generalized SIR model, which links the number of susceptible (and unprotected) people S, infectious (infected and spreading the infection) I and removed R (immunized, isolated and dead) over time t, [7,8]. Then the sum $V(t)=I(t)+R(t)$ is the theoretical estimation of the accumulated numbers of cases V_j and is derivative dV/dt is the estimation of the daily numbers of new cases.The values V_j corresponding to the period November 15–28, 2022 were taken in [9] for calculations the optimal SIR parameters according to the method presented in [8]. Since this period was before the beginning of easing the Zero-Covid restrictions in mainland China (November 30, 2022, [6]), the obtained forecast allows us to follow what the epidemic dynamics could be without the easing of the restrictions. Corresponding SIR curves are presented in Fig.1.

DOI: 10.54103/2282-0930/20637
Accepted: 31th January 2023

Fig.1.The results of the SIR simulations of the CoVID-19 pandemic wave in mainland China in late 2022, [9] and comparisons with the registered values and other calculations (markers).

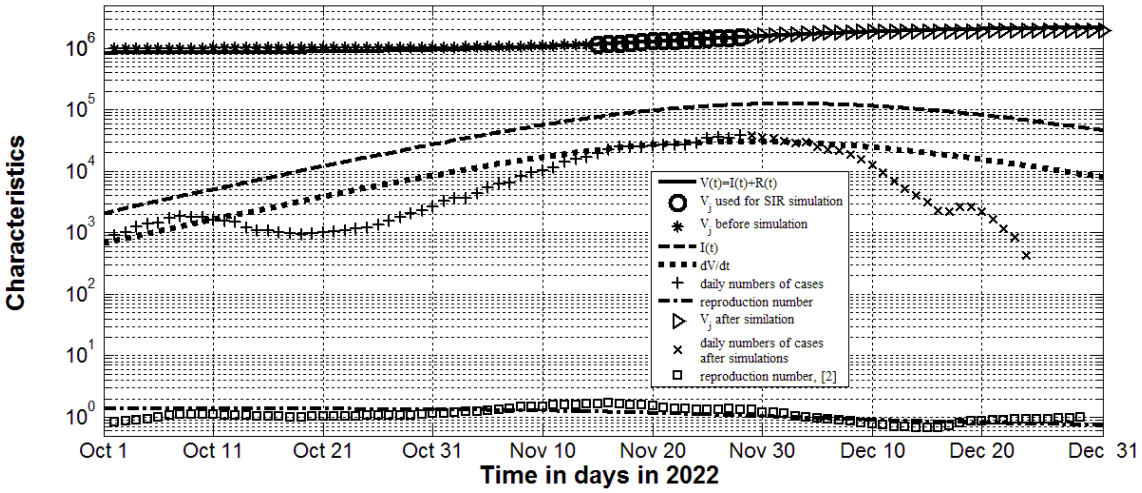


Table1. Cumulative numbers of laboratory-confirmed Covid-19 cases in mainland China for the period of October 1 to December 31, 2022 according to JHU report on January 4, 2023, [2].

Day in corresponding month of 2022	Accumulated numbers of cases, *Vj*		
	October	November	December
1	992046	1035560	1635593
2	992802	1038942	1698032
3	993657	1042353	1702449
4	994570	1046258	1733432
5	995650	1053804	1763679
6	996989	1054353	1788791
7	998779	1060280	1814055
8	1001871	1068278	1835148
9	1002305	1083133	1851983
10	1004398	1084571	1864738
11	1007872	1095429	1875371
12	1008241	1119690	1884218
13	1009732	1121325	1891352
14	1011191	1136846	1891352
15	1012384	1154441	1891352
16	1013390	1194415	1891352
17	1014291	1196687	1897331
18	1015212	1220590	1899290
19	1016148	1266052	1903956
20	1017089	1292056	1903961
21	1018065	1319652	1909905
22	1019038	1321605	1909905
23	1020028	1377221	1909905
24	1021117	1380570	1909905
25	1022331	1412498	1909905
26	1023743	1446896	1909905
27	1024984	1485399	1909905
28	1026239	1524446	1909905
29	1027864	1595756	1909905
30	1029918	1600201	1909905
31	1032790	-	1909905

The solid line show the number of victims *V(t)=I(t)+R(t)*, the dotted line – the theoretical estimations of the daily numbers of new cases *dV/dt*, the dashed line – the numbers of infectious persons *I(t)*, the dashed-dotted curve represent the effective reproduction numbers, calculated according to [8]. "Crosses" show the averaged daily numbers of new cases calculated with the use *Vj* values listed in Table 1 and formulas from [7, 8]."Stars", "circles", and "triangles" show the accumulated numbers of cases before, during and after SIR simulations, respectively. "Squares" show the values of the effective reproduction number from JHU dataset [2].

The registered accumulated numbers of cases (shown by "stars", "circles", and "triangles") are in very good agreement with the theoretical solid line. The smoothed registered daily numbers of cases

(shown by "crosses") follow the theoretical dotted line. Therefore, the lifting of restrictions probably did not increase the number of new cases.The slightly lower numbers of daily cases (compare "crosses x" with the dotted theoretical curve) can be explained by a decrease in the level of testing and the fact that many cases were not registered.

The generalized SIR model allows estimating the effective reproduction number, which shows the average number of people infected by one person [8,10-14]. The corresponding dashed-dotted line (see Fig.1, [9]) are close to the vales calculated with the use of method proposed in [14], listed in [2] and shown by "squares". The reproduction numbers do not exceed 1.7 and are much lower than value 21 reported by the director of National Institute For Viral Disease Control and Prevention, China CDC, [15]. Probably this incredible high value of the reproduction number has been used by NHC to estimate the recent epidemic dynamics in China.

The calculated numbers of infectious people *I(t)* make it possible to estimate the probability *p* of meeting a person spreading the infection, [7,8]. The maximum value 110,000 in early December, 2022 (see the dashed line) yielded the rather low value *p=7.6e-5* for mainland China, [9]. This value is much lover, than the maximum probability 0.012 estimated for Japan August 2022,[8].

The given analysis allows us to conclude that the information about the millions of new Covid-19 cases appearing every day in China is improbable.

ACKNOWLEDGEMENTS

The author is grateful to Oleksii Rodionov for his help in collecting and processing data.

REFERENCES

1. Bloomberg news. Internet information. https://www.bloomberg.com/news/articles/2022-12-23/china-estimates-covid-surge-is-infecting-37-million-people-a-day. Accessed January 3, 2023.
2. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://github.com/owid/covid-19-data/tree/master/public/data. Accessed January 3, 2023.
3. Ryan M Barber, Reed J D Sorensen, David M Pigott, Catherine Bisignano, Austin Carter, Joanne O Amlag et al. Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: a statistical analysis, The Lancet, Volume 399, Issue10344, 2022, pp.2351-2380, https://doi.org/10.1016/S0140-6736(22)00484-6.
4. I. Nesteruk, O. Rodionov. The COVID-19 pandemic in rich and poor countries. Preprint. Research Square. Posted December 7, 2022. https://doi.org/10.21203/rs.3.rs-2348206/v1.
5. Zero-COVID countries. Internet information https://en.wikipedia.org/wiki/Zero-COVID. Accessed January 3, 2023.
6. Internet information (in Chinese). https://news.cctv.com/2022/12/03/ARTIbCeNratl6uA4kXhzEdBG221203.shtml. Accessed January 3, 2023.
7. Nesteruk I. *COVID-19 pandemic dynamics*. Springer Nature. 2021. DOI: 10.1007/978-981-33-6416-5. https://link.springer.com/book/10.1007/978-981-33-6416-5.
8. Nesteruk I. Improvement of the software for modeling and forecasting the dynamics of epidemics and development of user-friendly interface. Preprint. Research gate, posted October 2022. DOI: 10.13140/RG.2.2.18942.66884.
9. Nesteruk I. The COVID-19 epidemic wave in mainland China at the end of 2022: monitoring and predicting with the use of the generalized SIR model. Preprint. Research Gate, posted December19, 2022. DOI:10.13140/RG.2.2.19479.24488.
10. Basic reproduction number. Internet information. https://en.wikipedia.org/wiki/Basic_reproduction_number.Accessed January 3, 2023.
11. Effective reproduction number. Internet information. https://www.r-bloggers.com/2020/04/effective-reproduction-number-estimation/. Accessed January 3, 2023.
12. an der Heiden, M., and O. Hamouda. 2020. "Schätzung Der Aktuel-Len Entwick lung Der Sars-Cov-2-Epidemie in DeutschLand – Nowcasting." *Epid-Bull* 17:10–15. https://doi.org/10.25646/669.
13. Cori, Anne, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. 2013. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics." *American Journal of Epidemiology* 178 (9):1505–12. https://doi.org/10.1093/aje/kwt133.
14. Arroyo-Marioli F, Bullano F, Kucinskas S, Rondón-Moreno C (2021) Tracking R of COVID-19: A new real-time estimation using the Kalman filter. PLoSONE 16(1):e0244474. https://doi.org/10.1371/journal.pone.0244474.
15. Internet information of National Institute For Viral Disease Control and Prevention, China CDC (in Chinese). https://news.cctv.com/2022/11/28/ARTI7xtooQZazf58JAzgVjoE221128.shtml. Accessed January 3, 2023.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

COVID-19 vaccination hesitancy and associated factors among the business community in Lira City, Uganda: a cross-sectional research

Eustes Kigongo[(1)], Everlyne Achan[(1)], Bosco Opio[(1)], Anne Ruth Akello[(1)], Marvin Musinguzi[(1)], Raymond Tumwesigye[(2)], Amir Kabunga[(3)], Sean Steven Puleh[(1)]

(1) Faculty of Public Health, Lira University
(2) Faculty of Nursing and Midwifery, Lira University
(3) Faculty of Medicine, Lira University

CORRESPONDING AUTHOR: Eustes Kigongo, Lira University, 1035, Lira City, Uganda, tel. +256770681031.
email: ekigongo@lirauni.c.ug

SUMMARY

Background: As of November 26, 2021, at least seven different vaccines across three platforms have been distributed globally. These vaccines include Pfizer/BioNTech released on December 31, 2020, AstraZeneca on February 16, 2021, Janssen by Johnson and Johnson on March 12, 2021, Moderna on April 30, 2021, Sinopharm on May 7, 2021, Sinovac CoronaVac on July 1, 2021, and COVAXIN on November 3, 2021. Despite this unprecedented scientific discovery, vaccine hesitancy is seen as a stumbling block towards achieving herd immunity in the battle to control this global pandemic. The effectiveness of vaccines has been based on the principle that the community was willing to take up the vaccine to achieve herd immunity. This study aimed to assess COVID-19 vaccination hesitancy and associated factors among the business community in Lira City, Uganda.
Methods: Descriptive cross-sectional design was conducted among the business community from Lira City in Northern Uganda. The sample size was 421, however, only 407 members of the business community who responded were included in the analysis.
Results: Of the 407 participants, 57.3% were females, 52.5% were married, 88.4% were Langi by Tribe, 43.5% had tertiary education, 33.4% were Anglicans, and 40.1% were market vendors. Results also show that about 32.3% of the respondents had either delayed or refused to take the COVID-19 vaccine. The correlates of vaccination hesitance were education level (aOR; 3.63, 95%CI; 1.49-8.79, p=0.04), having a chronic medical condition (aOR; 2.7, 95%CI; 1.39-5.38, p=0.04) and certainty in the COVID-19 vaccines (aOR; 0.27, 95%CI; 0.017-0.51, p=0.02). Respondents who had primary level education had a more than 2-fold increased odds of acceptance of COVID-19 vaccination compared to those who had not attained any formal education. Individuals who had chronic medical conditions had more than 2-fold increased odds of accepting the COVID-19 vaccine compared to those who did not have any chronic medical conditions. Those who were certain about the COVID-19 vaccine were 73% less likely to hesitate vaccination as compared to their counterparts who were uncertain.
Conclusion: The study found a substantially high level of COVID-19 vaccination hesitancy in Lira City and its predictors were level of education, chronic medical conditions and certainty in COVID-19 vaccines. For this reason, it is important to raise awareness among the business community about the vaccine. To increase uptake, policymakers and other stakeholders need to create effective communication techniques for behavior change.

Keywords: Acceptancy; COVID-19; Hesitancy; Vaccination; Vaccines.

DOI: 10.54103/2282-0930/20754
Accepted: 11th March 2023

BACKGROUND

The 2019 coronavirus disease (COVID-19) caused by the novel Coronavirus (SARS-CoV-2) began in the city of Wuhan in China and spread quickly across the world, generating a global health crisis of massive proportions [1]. The development of a vaccine against the virus is considered a pivotal moment in the efforts to curb disease spread and begin the resumption of normalcy in everyday life. As of November 26, 2021, at least seven different vaccines across three platforms have been distributed globally [2]. These vaccines include Pfizer/BioNTech released on December 31, 2020, AstraZeneca on February 16, 2021, Janssen by Johnson and Johnson on March 12, 2021, Moderna on April 30, 2021, Sinopharm on May 7, 2021, Sinovac CoronaVac on July 1, 2021, and COVAXIN on November 3, 2021[3]. Despite this unprecedented scientific discovery, vaccine hesitancy is seen as a stumbling block toward achieving herd immunity in the battle to control this global pandemic. The effectiveness of vaccines has been based on the principle that the community is willing to take up the vaccine to achieve herd immunity [4].

While the World Health Organization (WHO) targeted to provide at least 2 billion doses of the vaccine by the end of 2021,[2] there has been general apathy. Only 3% of the vulnerable population in Africa including health workers had been vaccinated by June 2021 [2]. The Pew Research Center reported on December 3rd, 2020 that 39% of Americans would probably, or definitely, not get a vaccine against COVID-19 [5]. In Uganda, the Ministry of Health planned to vaccinate 49.6% of the population by the end of 2021 but only 3.1% of the population has been fully vaccinated by 7th January 2022 [6]. Emerging studies show that attitude towards vaccines, low levels of health literacy, ill-health, lack of trust in the pharmaceutical industries producing the vaccines, gender, education level, age, and lack of knowledge were predictors of COVID-19 vaccination hesitancy [7], [8]. However, these factors are likely to differ from culture to culture and settings. A rapid survey of businesses in Uganda reveals that three-quarters of the surveyed businesses laid off employees due to the risks presented by COVID-19 and subsequent containment measures [9]. The results suggest that COVID-19 measures have reduced business activity by more than half [9]. Thus, it is important to understand the vaccination status of the business community in Uganda as livelihood depends on business continuity.

The COVID-19 pandemic has posed a substantial threat to the business community because it is a high-risk group due to its day-to-day interactions with the wider community [6]. They make cross-border movements and operated in settings with poor enforcement of standard operating procedures (SOPs), which practice threatens the success of the vaccination program. Thus, vaccination acceptance among the business community can reduce the threat. The delay in acceptance of the vaccine despite the availability of vaccination services is likely to increase transmission. The business community in Uganda in general and Lira City in particular in northern Uganda is not exceptional. Lira City is strategically located and serves as a business hub for the northern districts of Uganda. The district harbors people from all walks of life with various business ventures interacting with minimum adherence to the COVID-19 standard operating procedures (SOPs). A recent report in Western Uganda among the general population indicates that 53.6% were willing to accept vaccination [10]. However, few studies have been conducted among other high-risk groups including the business members in Uganda. Therefore, this study examined the level of COVID-19 vaccination hesitancy and its associated factors in the business community in Lira City.

METHODS

Study design and data collection

The study employed a cross-sectional design among 421 members of the business community from Lira City in northern Uganda. The study was conducted among the two divisions, East and West divisions of Lira city between November and December 2021. Data was collected using a questionnaire to measure the level of COVID-19 vaccination hesitancy and associated factors. Sampling was done and study participants were identified and approached to participate in the study after signing informed consent forms. Data were collected physically by five trained research assistants, and the process took around 20 minutes.

Study participants

A consecutive random sampling technique was employed to select a sample of 421 members of the business community including market vendors, retailers, and mobile money agents operating in Lira City. The sample size of the study was calculated using the Kish Leslie formula [11] for single proportions using a Z score of 1.96 at a 95% confidence interval, the level of COVID-19 vaccination hesitancy (p=46.4%) from a community study in south-western Uganda by [10], and an error margin of 5%, generating a sample size of 421 participants.

Ethical considerations

The present study was done in accordance with the Declaration of Helsinki. The study was also approved by Gulu University Research and Ethics Committee (GUREC-2021-115). Study participants were recruited based on written informed consent, and confidentiality

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

was maintained throughout the entire research process.

Study variables

Hesitancy to vaccinate was measured as a composite variable based on a six-items tool developed by the Strategic Advisory Group of Experts (SAGE) on a 5-point Likert scale [12]. The tool is based on the 3Cs model of complacency, confidence, and convenience of vaccines which are determinants of vaccination hesitancy [12]. Regarding the score of the tool, an average of the 6 items was obtained which generated a range of 1 to 5. These scores were categorized as hesitant and non-hesitant. An average score of ≤3.0 was considered non-hesitant and an average score of ≥3.0 was considered hesitant. The first section had demographic and factors associated with vaccine hesitancy. The Cronbach's alpha for the tool was 0.81. The study was conducted between November and December 2021. After approval of the study protocol, the city health authorities were approached and informed about the study.

Statistical analysis

The data were entered into a Microsoft Excel worksheet where pre-analysis cleaning was conducted. The Excel data was exported to STATA (Stata Corp LLC, TX, USA) version 15 software for analysis. Univariable, bivariable, and multivariable analyses were conducted. In univariable analysis, all variables were described and presented. Descriptive statistics including means, standard deviations, frequencies, and others were used to summarize the data. Univariate logistic regression was used to test the association at bivariable analysis and a p-value of less than 0.2 was considered significant. This was followed by binary logistic regression to identify predictors. This was performed at a 95% confidence interval and variables with a p-value of less than 0.05 were deemed significant.

RESULTS

Socio-demographic characteristics of study participants

Of the 421 respondents, only 407 responded generating a response rate of 96.7%. Therefore, only 407 responses were included in the final analysis. Table 1 shows that 57.3% of the respondents were females, 52.5% were married, 88.4% were Langi by tribe, 43.5% had tertiary education, 33.4% were Anglicans, and 40.1% were market vendors. Results in Table 1 also indicate that COVID-19 vaccination hesitancy was associated with chronic medical conditions (p=0.002), living in Lira City west (p<0.001), and single status (p=0.04).

Table 1: Socio-demographic characteristics of respondents (N=407)

Variable	Frequency (%)	COVID-19 vaccine hesitancy		P value
		Non-Hesitant n (%)	Hesitant n (%)	
Age category				
18-35	289(71.0)	196(71.0)	93(71.0)	Ref
36-65	114(28.0)	77(27.9)	37(28.2)	0.58
65 plus	4(1.0)	3(1.1)	1(0.8)	0.58
Gender				
Male	174(42.8)	118(42.8)	56(42.7)	Ref
Female	233(57.3)	158(57.2)	75(57.3)	0.53
Marital status*				
Separated or divorced	22(5.4)	15(5.5)	6(4.6)	Ref
Married	213(52.5)	144(52.6)	69(52.7)	0.18
Widowed	14(3.5)	9(3.3)	5(3.8)	0.67
Single	157(38.7)	106(38.7)	51(38.9)	**0.04**
Residence*				
City East	140(35.2)	95(35.2)	45(35.2)	Ref
City West	258(64.8)	175(64.8)	83(64.8)	**<0.001**
Level of education				
No education	47(11.6)	31(11.2)	16(12.2)	Ref
Primary	107(26.3)	73(26.4)	34(26.0)	**0.007**
Secondary	177(43.5)	120(43.5)	57(43.5)	0.92
Tertiary	76(18.7)	52(18.8)	24(18.3)	0.61
Have medical condition*				
No	343(84.5)	233(84.4)	110(84.6)	Ref
Yes	63(15.5)	43(15.6)	20(15.4)	**0.002**
Had COVID-19*				
No	356(87.9)	241(88.0)	115(87.8)	Ref
Yes	49(12.1)	33(12.0)	16(12.2)	**<0.001**

***has missing values; Ref=reference category**

COVID-19 Vaccination Hesitancy

Table 2 shows that out of the 407 respondents, only 131(32.2%) were hesitant against COVID-19 vaccination, 267(67.8%) were non-hesitant.

Table 2: COVID-19 Vaccination Hesitancy

COVID-19 Vaccination Hesitance	Frequency (n)	Percentage (%)
Hesitant	131	32.2
Non Hesitant	276	67.8
Total	**407**	**100**

Attitudes of respondents towards COVID-19 vaccination hesitancy

A bivariate analysis of attitudes about COVID-19 vaccination hesitancy was conducted. The COVID-19 vaccine is the best way to protect against COVID-19 ($p<0.001$) (Table 3).

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Table 3: Attitudes of respondents towards COVID-19 vaccination hesitancy (n=407)

Variable	COVID-19 vaccine hesitancy		P value
	Non-Hesitant n (%)	Hesitant n (%)	
Vaccine is safe			
No	99(71.4)	39(28.6)	Ref
Yes	177(66.0)	92(34.0)	0.35
Vaccine is essential			
No	87(75.3)	29(24.7)	Ref
Yes	189(65.1)	102(34.9)	0.09
The best way to protection			
No	157(80.4)	38(19.6)	Ref
Yes	122(57.3)	90(42.7)	**<0.001**
Awareness required			
No	13(64.3)	7(35.7)	Ref
Yes	264(68.2)	123(31.8)	0.76
COVID-19 causes serious complications			
Yes	35(67.6)	17(32.4)	Ref
No	241(67.9)	114(32.1)	0.97
COVID-19 is human-made			
No	129(68.5)	60(31.5)	Ref
Yes	156(71.4)	62(28.6)	0.59
Recommend vaccine			
No	52(60.3)	35(39.7)	Ref
Yes	222(69.4)	98(30.6)	0.19
Vaccine reduces transmission			
No	113(77.8)	32(22.1)	Ref
Yes	163(62.0)	99(38)	**0.006**
Vaccine was rushed			
No	155(66.0)	80(34.0)	Ref
Yes	123(71.5)	49(28.5)	0.31

Beliefs of respondents towards COVID-19 vaccination hesitancy

A bivariate analysis of beliefs about COVID-19 vaccination hesitancy was conducted. The vaccine reduces the risk of transmission (p=0.006), confidence in vaccine safety (p=0.05), and protection from hospitalization due to COVID-19 (p=0.01) were associated with COVID-19 vaccination hesitancy (Table 4).

Table 4: Beliefs of respondents towards COVID-19 vaccination hesitancy (n=407)

Variable	COVID-19 vaccination hesitancy		P value
	Non-Hesitant n (%)	Hesitant n (%)	
The vaccine protects from severe disease			
No	198(74.4)	98(25.6)	Ref
Yes	169(65.3)	89(34.7)	0.13
Vaccine prevents transmission			
No	93(71.0)	38(29.0)	Ref
Yes	184(66.5)	92(33.5)	0.45

Confidence in vaccine safety			
No	104(75.5)	34(24.6)	Ref
Yes	173(64.4)	96(35.6)	**0.05**
The vaccine can protect from COVID-19			
No	181(76.2)	57(23.8)	Ref
Yes	212(61.9)	130(38.1)	**0.01**
Vaccine protects hospitalization			
No	118(66.8)	56(33.2)	Ref
Yes	159(68.2)	74(31.8)	0.92
COVID-19 manufactured			
No	172(66.5)	86(33.5)	Ref
Yes	110(73.5)	39(26.5)	0.22
The vaccine helps those who already suffered from COVID-19			
No	172(68.2)	76(31.8)	Ref
Yes	106(68.5)	49(31.5)	0.95
Vaccine cause infertility			
No	237(71.4)	95(28.6)	Ref
Yes	44(59.1)	31(40.8)	0.09

Multivariate logistic regression analysis

In the multivariate analysis (Table 4), only factors that were significant at bivariate level with p≤0.2 were considered. The final model reported primary level of education (AOR; 3.63; 1.49-8.79; p=0.09), having a chronic medical condition AOR; 2.12; 1.016-4.44; p=0.04), and certainty in COVID-19 vaccine (AOR; 0.27; 0.017-0.51; p=0.02) as correlates of COVID-19 vaccination hesitancy.

Table 4: Correlates of COVID-19 vaccination hesitancy in the business community

COVID-19 vaccination hesitancy	AOR [95% confidence interval]	P value
Education level		
No education	Ref	
Primary level	3.63[1.49-8.79]	0.04*
Secondary level	0.90[0.54-3.00]	0.84
Tertiary level	1.38[0.44-4.34]	0.58
Have a chronic medical condition		
No	Ref	
Yes	2.70[1.39-5.38]	0.04*
Certainty in COVID-19 vaccine		
No	Ref	
Yes	0.27[0.017-0.51]	0.02*

*=Statistically significant attribute, Ref=reference category, AOR=adjusted odds ratio

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

DISCUSSION

We assessed factors associated with COVID-19 vaccination hesitancy among the business community in Lira City. About 32.3% had either delayed or refused to take the COVID-19 vaccine. Admittedly, the vaccine is the only effective intervention that prevents the death of people [13]. Our results indicate that a substantial number were hesitant. This may be attributed to the negative attitude towards the vaccine exhibited in the current study (Table 3), a point of view supported by other studies [14]. In addition, vaccine hesitancy is influenced by several factors including lack of confidence in the vaccine itself, lack of or misinformation about the vaccine, and a fear of side effects [15]. These findings are favorably comparable to the 30.7% hesitancy level observed among medical students in Uganda [16]. The differences in results may be attributed to the differences in attitude, sample size, settings, population, and time-lapse. These results imply more community education is needed and increase the uptake of the vaccine to achieve the herd immunity threshold of 80.3% [17].

Our results indicated that those with primary level education had a more than 3-fold increased odds of being hesitant to receive the COVID-19 vaccination compared to those who had not attained any formal education. Although this may be a surprising finding, other studies with a similar focus support it [18]. This result implies that those with primary education are a group who may be difficult to convince to get vaccinated, necessitating further vaccination efforts. The study's findings are exactly in contrast with US and German studies that found that low education level was a predictor of COVID-19 vaccination hesitancy [19], [20]. Similarly, a study in Southeast Asian countries indicated that a low level of education contributed to vaccination hesitancy [21].

Results in our study also show that individuals who had chronic medical conditions had a more than 2-fold increased odds of hesitating the COVID-19 vaccine compared to those who did not have any chronic medical conditions. Skepticism is more frequent among individuals with chronic diseases [22]. There is an assumption that vaccination, higher rates of side effects, and the interaction of vaccine and medication may worsen the condition [23]. Similar to our results, one study in Pakistan shows that chronic diseases are predictors of COVID-19 vaccination hesitancy [24].

Additionally, according to the results of our study, those who were certain about COVID-19 vaccines were 73% less likely to vaccinate as compared to their counterparts who were uncertain. Certainty in vaccines is key in decision-making and may be a great determinant in COVID-19 vaccine uptake [13], [25]. Uncertainty about the vaccine is likely to culminate in the non-acceptance of the COVID-19 vaccine. Studies have shown that attitude determine COVID-19 vaccine uptake [13], [25].

LIMITATIONS

The study has some limitations. The sample size is large but may not be adequate to generalize the study findings. The study also is limited by geography as it was only conducted in Lira City, therefore, a cross-national study is highly recommended. Lastly, the study was also cross-sectional so we cannot conclude causality.

CONCLUSION

The study reports a substantially high level of COVID-19 vaccination hesitancy in Lira City and its predictors were level of education, chronic medical condition, and certainty in COVID-19 vaccines. Therefore, there is a need to sensitize the business community about the vaccine to increase uptake. Policymakers and other stakeholders also need to develop effective behavior change communication strategies to improve uptake.

ACKNOWLEDGMENTS

The authors wish to acknowledge the participants in this study.

FUNDING

There is no funding source for this publication.

AUTHORS' CONTRIBUTIONS

EK, SSP, EA, and AK conceptualized and planned the study. EK and BO wrote the methodology and performed the statistical analysis. MM, RT and EA prepared questionnaires and conducted the survey. RT and ARA edited the manuscript for important intellectual content. AK drafted the manuscript for publication. All authors contributed to the article and gave final approval for publication.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

The study was also approved by Gulu University Research and Ethics Committee (GUREC-2021-115). Study participants were recruited based on written informed consent, and confidentiality was maintained throughout the entire research process. Study

participants were identified and approached to participate in the study after signing informed consent forms.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

1. H. Zhu, L. Wei, and P. Niu, "The novel coronavirus outbreak in Wuhan, China," *Global Health Research and Policy*, vol. 5, no. 1, p. 6, Mar. 2020, doi: 10.1186/s41256-020-00135-6.
2. WHO, "COVID-19 vaccines | WHO | Regional Office for Africa," 2021. https://www.afro.who.int/health-topics/coronavirus-covid-19/vaccines (accessed Jul. 26, 2021).
3. WHO, "COVID-19 vaccines," 2021. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines (accessed Jun. 20, 2021).
4. F. E. Andre *et al.*, "Vaccination greatly reduces disease, disability, death and inequity worldwide," *Bulletin of the World Health Organization*, vol. 86, no. 2, pp. 140–146, 2008, doi: 10.2471/BLT.07.040089.
5. C. Funk and A. Tyson, "Intent to Get a COVID-19 Vaccine Rises to 60% as Confidence in Research and Development Process Increases," *Pew Research Center*, no. December, pp. 1–29, 2020.
6. MoH, "Uganda receives 864,000 doses of COVID-19 vaccines - Uganda | ReliefWeb," 2021. https://reliefweb.int/report/uganda/uganda-receives-864000-doses-covid-19-vaccines (accessed Jul. 26, 2021).
7. R. H. El-Sokkary *et al.*, "Predictors of COVID-19 vaccine hesitancy among Egyptian healthcare workers: a cross-sectional study," *BMC infectious diseases*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/S12879-021-06392-1.
8. E. Robertson *et al.*, "Predictors of COVID-19 vaccine hesitancy in the UK household longitudinal study," *Brain, Behavior, and Immunity*, vol. 94, pp. 41–50, May 2021, doi: 10.1016/J.BBI.2021.03.008.
9. P. C. Lakuma, N. Sunday, B. Sserunjogi, R. Kahunde, and E. F. Munyambonera, "How has the COVID-19 pandemic impacted Ugandan businesses? Results from a business climate survey," 2020.
10. I. Echoru, P. D. Ajambo, E. Keirania, and E. E. M. Bukenya, "Sociodemographic factors associated with acceptance of COVID-19 vaccine and clinical trials in Uganda: a cross-sectional study in western Uganda," *BMC Public Health*, vol. 21, no. 1, p. 1106, Dec. 2021, doi: 10.1186/s12889-021-11197-7.
11. A. S. Singh and M. B. Masuku, "Sampling techniques & determination of sample size in applied statistics research: An overview," *International Journal of economics, commerce and management*, vol. 2, no. 11, pp. 1–22, 2014.
12. N. E. MacDonald *et al.*, "Vaccine hesitancy: Definition, scope and determinants," *Vaccine*, vol. 33, no. 34, pp. 4161–4164, Aug. 2015, doi: 10.1016/J.VACCINE.2015.04.036.
13. K. A. Fisher, S. J. Bloomstone, J. Walder, S. Crawford, H. Fouayzi, and K. M. Mazor, "Attitudes Toward a Potential SARS-CoV-2 Vaccine : A Survey of U.S. Adults," *Annals of internal medicine*, vol. 173, no. 12, pp. 964–973, Dec. 2020, doi: 10.7326/M20-3569.
14. G. Huynh, T. Van Nguyen, D. D. Nguyen, Q. M. Lam, T. N. Pham, and H. T. N. Nguyen, "Knowledge About COVID-19, Beliefs and Vaccination Acceptance Against COVID-19 Among High-Risk People in Ho Chi Minh City, Vietnam," *Infection and Drug Resistance*, vol. Volume 14, pp. 1773–1780, May 2021, doi: 10.2147/idr.s308446.
15. U. Samarasekera, "Feelings towards COVID-19 vaccination in Africa," *The Lancet. Infectious diseases*, vol. 21, no. 3, p. 324, Mar. 2021, doi: 10.1016/S1473-3099(21)00082-7.
16. A. M. Kanyike *et al.*, "Acceptance of the coronavirus disease-2019 vaccine among medical students in Uganda," *Tropical Medicine and Health*, vol. 49, no. 1, pp. 1–11, Dec. 2021, doi: 10.1186/s41182-021-00331-1.
17. WHO, "Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19," 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/herd-immunity-lockdowns-and-covid-19?gclid=Cj0KCQjw_viWBhD8ARIsAH1mCd6Jle2BJPiz-RPtFiYPG6T9GEf8rksOwOHrVHf-hlyatSt6YD8cutQaAsgAEALw_wcB (accessed Jul. 07, 2022).
18. V. Sypsa, S. Roussos, V. Engeli, D. Paraskevis, S. Tsiodras, and A. Hatzakis, "Trends in COVID-19 Vaccination Intent, Determinants and Reasons for Vaccine Hesitancy: Results from Repeated Cross-Sectional Surveys in the Adult General Population of Greece during November 2020–June 2021," *Vaccines*, vol. 10, no. 3, Art. no. 3, Mar. 2022, doi: 10.3390/vaccines10030470.
19. V. M. Beusekom, "Lack of high school education predicts vaccine hesitancy | CIDRAP," pp. 1–3, 2022.
20. A. Zychlinsky Scharff *et al.*, "Students' age and parental level of education influence COVID-19 vaccination hesitancy," *European Journal of Pediatrics*, pp. 10–15, 2021, doi: 10.1007/s00431-021-04343-1.
21. R. R. Marzo *et al.*, "Hesitancy in COVID-19 vaccine uptake and its associated factors among the general adult population: a cross-sectional study in six Southeast Asian countries," *Tropical Medicine and Health*, vol. 50, no. 1, pp. 1–10, 2022, doi: 10.1186/s41182-021-00393-1.
22. M. Rakusa *et al.*, "COVID-19 vaccination hesitancy among people with chronic neurological disorders: A position paper," *European Journal of Neurology*, vol. 29, no. 8, pp. 2163–2172, 2022, doi:

10.1111/ene.15368.
23. K. Yamamoto, "Adverse effects of COVID-19 vaccines and measures to prevent them," *Virology Journal*, vol. 19, no. 1, p. 100, Jun. 2022, doi: 10.1186/s12985-022-01831-0.
24. A. Malik, J. Malik, and U. Ishaq, "Acceptance of COVID-19 vaccine in Pakistan among health care workers," *PLoS ONE*, vol. 16, no. 9 September, pp. 1–11, 2021, doi: 10.1371/journal.pone.0257237.
25. C. Al Halabi, S. Obeid, H. Sacre, M. Akel, R. H.-B. P. Health, and undefined 2021, "Attitudes of Lebanese adults regarding COVID-19 vaccination," *Springer*, 2020.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Using accelerometer Analysis to Assess Physical Activity and Sedentary Behavior in Syrian Adults

Mahfouz Al-Bachir[(1)], Husam Ahmad[(1)]

(1) Department of Radiation Technology, Atomic Energy Commission of Syria.

CORRESPONDING AUTHOR: Dr. Mahfouz Al-Bachir, Department of Radiation Technology, Atomic Energy Commission of Syria, P.O. Box 6091, Damascus, Syria. Email: ascientific9@aec.org.sy

SUMMARY

Background: Overweight and obesity has become a serious public health problem in the world. Changes in daily physical activity (PA) levels may help moderate the health risks of sedentary behavior (SB). The aim of the present study was to determine PA and SB by triaxial accelerometer (ActiGraph GTX3 GT3X+, ActiGraph, Pensacola, FL. 32502 USA) in Syrian adults that affected by age, marital, body mass index (BMI), education, and smoking status.

Methods: We used cross-sectional baseline data from 97 participants of adults (18-60 years). Subjects wore an accelerometer for 7 consecutive days. Magnitude counts/minute was extracted to determine time in inactivity, in low-intensity, moderate, and vigorous-to-very-vigorous activity.

Results: Higher age was associated with more time in all categories of PA (light, moderate vigorous MVPA with bouts, MVPA without bouts, and step per day), while higher BMI was related to less time in PA, overweight/obese subjects, on daily average, spent less, but not significantly, time standing and little more time spent in sitting than the normal weight groups. Participant comparing to non-smoking peoples tended to spent less time in PA; however, this synchronization was not considerable.

Conclusion: Finally, our investigation demonstrated a positive synchronization between sedentary time and educational level. In this cohort of adults, most of men and women fulfilled the WHO recommendations. The levels of PA in 18–60-year-old adults are similar to previous data reported in adults.

Keywords: accelerometer; physical activity; sedentary behavior; Syria.

INTRODUCTION

Several urgent calls to action have been done to combat that the international physical activity (PA) and sedentary behavior (SB) [1]. Low PA is associated with an increased risk of morbidity and mortality. While, participation in regularly PA is well documented in the public health benefits across the life course [2].

Recent epidemiological studies have assessed the relationships between health risk associated with overweight or obesity and sedentary behavior (SB) status [3]. Regular physical activity (PA) is associated with good health [4]. In particular, SB and a lack of PA has been found to be associated with numerous health diseases, and considered to be an important burden in public health and is a health risk factors [5,6]. Published report classified the impact of PA as similar to that of smoking in relation to the risk of non-communicable diseases globally [7].

According to the WHO guidelines on PA and SB; the recommended amount of PA that adults should engage in at least 150-300 minutes per week of moderate PA, at least 75-150 min per week of vigorous PA, or an equivalent combination of the two recommendations listed above to prevent non-communicable diseases [8].

Accurate measurement of PA patterns (duration, frequency and intensity) is demand for effective intervention of non-communicable chronic diseases prevention programmers [4]. PA is difficult to determine

DOI: 10.54103/2282-0930/20755
Accepted: 21th March 2023

in population based health survey. Presently, several objective and subjective self report instruments are available for measuring PA in the population [9]. Objective assessments, such as pedometer, that count steps, and accelerometers, that determine movement intensity have been commonly used technique to measure PA levels has been successfully used in large scale epidemiological studies in different human generations [10,11]. To date, the majority of researches has focused on PA towards in developed countries and has indicated in the literature gape concerning PA in developing countries, where the type of activities performed differ from those taking place in high income settings [12].

Equipment's have been improved and procedures have been used to report situations in previous scientific research. To date, no studies have examined the Syrian situations using these devices. There is currently limited or no information about the feasibility and precision in Syria. However, there is abundant literature on the advanced analytical methods to assess PA behavior using accelerometer time series on other populations [13]. This information would be particularly useful for clinicians and researchers when deciding this instrument to chose. Therefore, The objective of this work was to determine the synchronization of SB and intensity of PA with age, marital status, BMI, education level and smoking among adult men and women using objectively assessed data on PA and SB in Syrian adults.

MATERIAL AND METHODS

Study design and sample recruitment

This was an observational cross sectional designed study. During 2020-2021, a sample of 121 participants were randomly selected from several workplaces in Syrian Atomic Energy Commission (SAEC). In total, 97 had valid accelerometer measurements. Participants were required to be between the ages of 18 and 60 years. Participants were eligible for this study if they were willing to wear AG accelerometer for 7 consecutive days and were willing to complete surveys in Arabic language. Each participant provided informed consent prior to participation after a detailed explanation of the study protocol. The Atomic Energy human ethics committee approved the study protocol. It was excused in accordance with the regulation prescribed by Helsinki Declaration of the world Medical Association.

All subjects with complete data on objectively assessed PA, height and weight were included in the current analyses. Height and weight in light clothing were determined to the nearest 0.5 cm with a wall-mounted stadiometer (Seca, Model: 225 1721009; Germany), and to the nearest 0.1 kg with electronic digital scales (Seca, Model: 7671321004; Germany), respectively. From these measurements the body mass index (BMI) was calculated as weight (kg) divided by height (m) squared (BMI, in kg/ m^2).

Accelerometer processing

A triaxial accelerometer (ActiGraph GTX3 GT3X+, ActiGraph, Pensacola, FL. 32502 USA) was used to assess PA for the participants with four to seven days in this study. The device was initialized to collect data on PA, including activity counts, energy expenditure (kcal), steps, and activity intensity as metabolic equivalents (METs) [5]. Participants were advised to wear the accelerometer on the right left hip for seven consecutive days during waking hours excluding contact sports, washing, bathing, swimming or sleeping activities [14]. Subjects were asked to remove the device before aquatic activities such as showering, swimming or bathing. The AG accelerometer data was processed using AG ActiLife 6 software, and exported to Microsoft Excel format. Within Microsoft Excel, minutes of PA, including light, moderate and vigorous and SB on valid days (≥10 hours of wear time) were summed and divided by the number of valid days of wear time to create a daily average equivalent activity. ActiLife 6 software to initialize the accelerometer and to download results, raw data was converted with Freedson cut points [15]. PA intensity levels refers to how hard your body is working during PA, and defined as that person's total energy expenditure (TEE) in a 24-hour period by his or her basal metabolic rate (BMR). Average daily time in moderate to vigorous physical activity (MVPA) (min/day) and SB (min/day) were calculated [16]. The daily average was multiplied by seven to create a weekly total [9].

Statistical analysis

Differences in PA and SB levels between adult's women and men were analyzed using Student's t-test, and differences according to the ages, marital status, BMI, education level, and smoking statues were analyzed using paired-samples t-test using Fisher's one-way ANOVA. In cases of unequal variances, Welch's ANOVA was used. All analyses were performed using Statistical Package of Social Science (SPSS) (Version 17.0.1, 2001 24, 2016, SPSS Inc., Chicago, USA) and all values of $p < 0.05$ were considered statistically significant.

RESULTS

Participant characteristics

The overall and sex specific characteristics of the participants (N = 97) are indicated in Table 1. Most of participants were women (n= 67, 69.1%). The average

age (SD) of participants ranged from 36.4 (±8.7) years among women to 44.3 (±6.9) years among men. A total of 97 individuals (30.9% males; 69.1% females) with a mean age of 38.9 years (SD = 8.9) participated in the study. Out of 97 participants, 25 (25.8%) were singles, and 72 (74.2%) were married. The mean BMI was 27.6 kg/m2 (SD = 4.7), BMI = 28.8±3.3 kg/m^2 for men, and 27.1±5.2 kg/m2 for women. The majority of the subjects were overweight, as 69.1% (n=67) were classified as overweight or obese. As for education status, 12(12.4%) were lower education (< secondary school), 17(17.5%) were moderately at education (secondary school), while 68 (70.1%) declared having a higher education levels (> secondary school).

Overall physical activity

Means and standard deviations for the accelerometer variables are presented in Table 2. The participants accumulated a total average daily sedentary behavior, light PA, moderate PA, and Vigorous PA time of 644.2 min/day (10.74 hours), 196.0 min/day (3.27 hours), 41.0 min/day (0.68 hours), 1.0 min/day (0.02 hours), respectively. The MVPA is 7.3 min/day (0.12 hours) with 10 min bouts, and 42.0 min/day (0.70 hours) without 10 min bouts. Participants accumulated a mean of 7502.2 steps per day. There were only slight, non-significant differences between male and female in time spend sedentary or in light PA (Table 2). Men accumulate more minutes of moderate, vigorous, MVPA with 10 min and , MVPA without 10 min bouts than women ($p<0.05$). Also, men accumulate more steps per day than women ($p<0.001$).

Physical activity levels by age, marital status, BMI, educational level and smoking status

All physical activity measures were positively associated with age. While sedentary behavior was inversely associated with age. The associations were significant ($p<0.05$) only for light and vigorous PA (Table 2).

Sedentary behavior, light PA, Moderate PA, Vigorous PA, MVPA with bouts and MVPA without bouts were not associated with marital status, BMI, education level and smoking ($p>0.05$) (Table 2).

Step per day were inversely synchronization with education levels ($P<0.05$). There were no differences in accumulated step per day between ages marital status, BMI values, and smoking (Table 2).

Meeting recommendation

According to the global WHO recommendations for PA, 88.7%. of adults (95% CI: 80.6 – 94.2%) (85.1% of women (95% CI: 82.8 – 99.9%) and 96.7% of men (95% CI: 74.3 – 92.6%)) accumulate 150 minutes per week of MVPA without 10 minute bouts, while 3.1% of adults (95% CI: 0.6 – 8.8%) (1.5% of women (95% CI: 0.8 – 22.1%) and 6.7% of men (95% CI: 0.0 – 8.0%)) accumulate 150 minutes per week of MVPA with 10 minute bouts. The same trend was indicated for the age groups (Figure 1).

According to the proportion fulfilling the global WHO recommendations for step count of 10.000 steps per day, 88.7% of adults accumulate < 10.000 step per day (73.3% for men and 95.5% for women), and 11.3% of adults accumulate > 10.000 step per day (26.7% for men and 4.5% for women). The same trend was indicated for the age groups (Figure 2).

DISCUSSION

This study reports cross-sectional associations between PA and SB that measured by objective method which the accelerometer, in a group of adult subjects of both sexes age 18–60 yr. The accelerometer data are the first objective measurements of PA and SB in a nationally representative survey. However, the absolute count, duration results from the accelerometer data provide a new and a real picture of PA in the Syrian population. Accelerometers are widely used for estimation determining PA in free living conditions. The accelerometer provides an estimate close to truth, and that respondents greatly overestimate their PA [17].

The World Health Organization (WHO) recommendation is to practicing at least 150 min per week of moderate-intensity or 75 min per week of vigorous-intensity aerobic PA per week, or an equivalent combination of both type of PA, for people aged from 18 to 64 years [8,18]. National and international guidelines recommended 30 minutes or more of moderate intensity PA daily (at least five day a week), or vigorous intensity PA for at least 20 minutes and three day a week [19]. However, in assessed Syrian group the moderate intensity PA is high (40 min per day) While, the vigorous intensity PA is low (only 1 min per day). The lack of PA may contribute to the deteriorations in health observed among this population.

The most important finding in this study is that 88.7% of Syrian adults are meeting the normal PA recommendation (accumulate 150 minutes per week of MVPA without 10 minute bouts). When activity in bouts of 10 minutes was considered, adherence prevalence estimates were 3.1% among adults. A considerable amount of the evidence in support of the 150-minutes-per-week recommendation suggests that frequent PA is important for health [19]. Recently, based on data from accelerometer, the revised United States recommendations for PA omitted the requirement that MVPA should be performed in at least 10-minute bouts [20].

Our results, in agreement with previous studies, may support the opinion that meeting the recommended PA could may not be sufficient if sedentary time is not

reduced [21, 22]. It has been recommended that SP SB could be an independent determent of health risk [23].

Objectively measured PA data show that 3% of Americans aged 20 to 59 years accumulating at least 30 minutes of MVPA in 10 minute bouts on 5 out of 7 days [24]. Canadian date, also show that 6% for the same age range accumulating at least 150 minutes a week of MVPA in 10 minute bouts [25]. The low prevalence of adherence to the WHO PA suggestion is comparable to other European studies. In Germany only a median of 14% of MVPA was aggregated in bouts of at least 10 minutes for the testing group (N = 475), aged 48-68 years [24]. In Norway, 20% of the study participants aged 20±85 years, aggregated at least 30 minutes of daily MVPA in bouts of 10 minutes [26]. In Swedish subjects (age range 18-79 years), only 1% reached 30 minutes/day in pout of 10 minute bouts [28]. In Portugal, 3-7% participants aged 40-64 years aggregated at least 30 minutes MVPA/day in periods of at least 10 minutes [29]. In Norwegian adults and elderly, 22% fulfilled the current global recommendation for PA. However, when counting all accumulated non-bouted MVPA, the proportion increased three-fold, to 70% [30].

Our results show that 11.3% of Syrian adults (men and women) achieved the 10.000 steps per day. The average adults ranged from 6.9661 to 9.277 steps per day. These figures are close to other results, which showed that American adults accumulated about 9,700 steps per day [24]. For Canadian adults the average man takes approximately 9,500 steps per day, and the average woman, 8,400 steps. [25].

Males are generally more active than females, and PA is higher in older age groups. However, the mean count, duration, and adherence prevalence results from the accelerometer data provide a new picture of PA and SB in the Syrian population; by ages 18–60 yr, mean levels of moderate activity are low, and vigorous activity is almost nonexistent. The light levels of PA are particularly evident when bouts of activity are considered. Our findings suggest that men and women spend same time in what might be considered low or sedentary levels of PA, while men spend more time in moderate and vigorous levels of activity. In previous experiments, male subjects participants in studies from Germany [26], Norway [31], Portugal [32], the United States [32], and the United Kingdom [33] aggregated more minutes of MVPA than female participants.

With respect to age, we observed small differences in the summary measure of PA by age. Higher age was related to more time in light, moderate vigorous MVPA with bouts, MVPA without bouts, and step per day. These differences by age group are in contrast to other population-based studies [34] that have used hip worn accelerometers.

Unexpectedly, our investigation found no significant difference in sedentary time and PA (light, moderate, vigorous, MVPA with bouts, MVPA without bouts, and step per day) between normal weight and overweight/obese groups. However, the results of this study support the hypothesis that obese or overweight participants, on daily average, spent less, but not significantly, time standing as MVPA, and little more time sitting than the normal weight groups. These results are comparable to the study conducted by Schaller et al. [35], who also found no significant difference in standing and sitting time between normal and overweight participants. Jaeschke et al. [36], who did not find as association of BMI with other activity intensities. Previous study illustrated that normal and obese participants spent approximately the same amount of time for lying down [37]. Several studies support the hypothesis that obese participants spent less standing time and more sitting time than normal or overweight groups [3]. Additionally, higher BMI was associated with less time spent in low intensity activity [38], which may be explained by the fact that BMI is strongly correlated with body weight [35]. The obese individuals spending more sitting time (using the computer or watching TV) compared to normal weight or overweight groups [36]. The relative instability of SB and PA time in-between normal and obese participants need more studies for confirmation.

Our analyses suggested differences by education levels when assessing by accelerometer. Our study demonstrated a positive association between sedentary time and educational level, which is consistent with studies from other countries [39]. Also, our results demonstrated a negative association between PA and educational level, which is in contrast with studies from other countries [40,41]. Individuals with lower education are more likely to possess jobs including standing and walking, usually of light intensity PA [42].

Current smokers spent less time in moderate and vigorous activity and more time in low intensity [43]. Additionally, smoking is often related to a generally less healthy attend lifestyle including lower sports and exercises behavior [44]. Thus, one may conclude that smoking has a long-term effect on intensity of individual PA. This assumption is supported by our founding that participant compared to never smokers tended to spent less time in PA; however, this association was not statistically significant.

Strengths of the study

This study had several strengths, beginning with the adherence to standardized WHO protocols in administering accelerometer, and the concordant measurement period (7days) for the accelerometer employ standardized WHO [8] protocols in estimating PA and the concordant measurement period (7days) for the accelerometer. Strength of this study was the focus on participants under free-living conditions. Also, our study has strength that we objectively measured PA levels and SB time using a triaxial accelerometer for the first time in Syria.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Limitations of the study

We only included wake time data; therefore, limitation was that sleep period was excluded from the data collected and processes. New findings supports that sleep period and sleep disorders may negatively affect of health outcomes [45]. The study population, consisting of predominantly participants of higher socioeconomic situations and healthy adults, is not representative of the general population. It is not possible to confirm that these results are representative of the wider male and female population residing in Syria. Further works are required using larger samples of population from multiple Syrian regions.

Moreover, the accelerometer cut-points for categorizing the intensity of PA may be population-specific and not appropriate for Syrians. However, the classification of PA intensities by using accelerometer information is also affected on the cut points used [46].

CONCLUSION

This study is one of the first to use the accelerometer function-measured time spent in standing, and sitting postures among diverse samples of adults from different gander, wide age range, different education statues, different BMI, and smoking or not smoking participants of Syrian adults, showed moderate validity of the long IPAQ when compared to accelerometer data with correlations in a similar range as reported in other studies.

CONFLICT OF INTEREST

The authors declare no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGMENTS

The authors wish to express their deep appreciation to the Director General of AECS.

REFERENCES

1. Cleland C, Ferguson S, Ellis G, Hunter RF. Validity of the International Physical Activity Questionnaire (IPAQ) for assessing moderate-to-vigorous physical activity and sedentary behaviour of older adults in the United Kingdom. BMC Medical Research Methodology, 2018; 18, 176 https://doi.org/10.1186/s12874-018-0642-3.
2. Carlin A, Perchoux C, Puggina A, Aleksovska K, Buck C, Burns C, et al. A life course examination of the physical environmental determinants of physical activity behaviour: A Determinants of Diet and Physical Activity (DEDIPAC) umbrella systematic literature review. PLoS ONE.2017; 12(8): e0182083. https://doi.org/ 10.1371/journal.pone.0182083.
3. Barwais FA, Cuddihy TF, Rachele JN, Washington TL. ActiGraph GT3X determined variations in free living standing, lying, and sitting duration among sedentary adults. Journal of Sport and Health Science.2013; 2: 249-256.
4. Agogo GO, Voet H, Hulshof PJM, Veer P, Trijsburg L, Eeuwijk FA, Boshuizen HB. Validation of accelerometer for measuring physical activity in free-living individuals. Baltic Journal of Health and Physical Activity.2018; 10 (1): 7-21.
5. Yano S, Koohsari MJ, Shibata A, Ishii K, Mavoa S, Oka K. Assessing Physical Activity and Sedentary Behavior under Free-Living Conditions: Comparison of Active Style Pro HJA-350IT and ActiGraphTM GT3X+. Int J Environ Res Public Health.2019; 16(17): 3065. Doi: 10.3390/ijerph16173065.
6. Leinonen AM, Ahola R, Kulmala J, Hakonen H, Vaha-Ypya H, Herzig KH, Auvinen J, Kiukaanniemi SK, Sievanen H, Tammelin TH, Korpelainen R, Jamsa T. Measuring physical activity in free-living conditions. Comparison of three accelerometer-based methods. Front Physio.2017; 10 January.Doi: 10.3389/fphs.2016.00681.
7. Lee M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012 Jul 21;380(9838): 219-229. doi: 10.1016/S0140-6736(12)61031-9.
8. World Health Organization (WHO) 2020. WHO guidline on physical activity and sedentary behavior, 25 November 2020. Sports Medicine volume 54 issue 24. http://www.researchgate.net.
9. Oyeyemi AL, Umar M, Oguche F, Aliyu SU, Oyeyemi AY. Accelerometer-Determined Physical Activity and Its Comparison with the International Physical Activity Questionnaire in a Sample of Nigerian Adults. PLoS ONE.2014; 9(1): e87233. doi:10.1371/journal.pone.0087233.
10. Plasqui G, Bonomi AG, Westerterp KR. Daily physical activity assessment with accelerometers: New insights and validation studies. Obes.2013; Rev. 14(6): 451-462.
11. Nicaisa V, Crespo NC, Marshall S. Agreement between the IPAQ and accelerometer for detecting intervention related changes in physical activity in sample of Latin women. Journal of Physical Activity and Health.2014; 11: 846-852.
12. Hallal PC, Reichert FF, Clark VL, Cordeira KL, Menezes MB, Eaton S, Ekelund U, Wells JC. Energy expenditure compared to physical activity measured by accelerometry and self-report in adolescents: A validation study. PloS ONE.2013; (11): e77036, doi:10.1371/journal.pone.0077036.
13. Backes A, Gupta T, Schmitz S, Fagherazzi G, Hees V, Malisoux L. Advanced analytical methods to assess physical activity behavior using accelerometer time

series: A scoping review. Scand J Med Sci Sports, 2022; 32: 18-44. DOI: 10.1111lsms.14085.
14. Murphy JJ, Murphy MH, MacDonncha C, Murphy N, Nevill AM, Woods CB. Validity and Reliability of Three Self-Report Instruments for Assessing Attainment of Physical Activity Guidelines in University Students. Measurement in Physical Education and Exercise Science.2017; 21(4): 1-8. http://dx.doi.org/10.1080/1091367X.2017.1297711.
15. Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications, Inc accelerometer. Med Sci Sports Exerc.1998; 30(5): 777–781.
16. Aadland E, Ylvisåker E. Reliability of the Actigraph GT3X+ accelerometer in adults under free-living conditions. PLoS One.2015; 10(8): e0134606.
17. Duncan GE, Sydeman SJ, Perri MG, Limacher MC, Martin AD. Can sedentary adults accurately recall the intensity of their physical activity? Prev Med.2001; 33: 18–26.
18. Helo K, El Helou N, Mahfouz M, Mahfouz Y, Salameh P, Harmouche-Karaki M. Validity and reliability of an adapted arabic version of the long international physical activity questionnaire. BMC Public Health.2018; 18: 49. DOI 10.1186/s12889-017-4599-7
19. Nicolaou M, Gademan MGJ, Snijder MB, Engelbert RHH, Dijkshoorn H, Terwee B, Stronks K. Validation of the SQUASH physical activity questionnaire in a multi-ethnic population: The HELIUS study. PLOS.2016; August 30: 1-14.
20. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, George SM, Olson RD. The physical activity guidelines for Americans. JAMA.2018; 320(19): 2020–2028. https://doi.org/10.1001/jama.2018.14854 PMID: 30418471.
21. González K, Fuentes J, Márquez JL. Physical inactivity, sedentary behavior and chronic diseases. Korean J Fam Med.2017; 38(3): 111–115. doi:10.4082/kjfm.2017.38.3.111
22. Aljuhani O, Alkahtani S, Alhussain M, Smith L, Habib SS. Associations of Physical Activity and Sedentary Time with Metabolic Syndrome in Saudi Adult Males. Risk Management and Healthcare Policy.2020; 13: 1839–1847
23. Edwardson CL, Gorely T, Davies MJ, Gray LJ, Khunti K, Wilmot EG, Yates T, Biddle SJH. Association of sedentary behaviour with metabolic syndrome: a meta-analysis. PLoS One.2012; 7(4): e34916. doi:10.1371/journal.pone.0034916.
24. Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-determined steps per day in US adults. Medicine and Science in Sports and Exercise.2009; 41(7): 1384-1391.
25. Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: Accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health Reports.2011; 22(1): 1-8.
26. Luzak A, Heier M, Thorand B, Laxy M, Nowak D, Peters A, Schulz H. Physical activity levels, duration pattern and adherence to WHO recommendations in German adults. PLoS ONE.2017; 12(2): e0172503. doi:10.1371/journal. pone.0172503.
27. Hansen BH, Kolle E, Dyrstad SM, Holme I, Anderssen SA. Accelerometer-determined physical activity in adults and older people. Med Sci Sports Exerc.2012; 44(2): 266-272. Epub 2011/07/29. doi: 10.1249/ MSS.0b013e31822cb354 PMID: 21796052
28. Hagstromer M, Oja P, Sjostrom M. Physical activity and inactivity in an adult population assessed by accelerometry. Med Sci Sports Exerc.2007; 39(9): 1502-1508. Epub 2007/09/07. doi: 10.1249/mss. 0b013e3180a76de5 PMID: 17805081
29. Baptista F, Santos DA, Silva AM, Mota J, Santos R, Vale S, Ferreira JP, Raimundo AM, Moreira H, Sardinha LB. Prevalence of the Portuguese population attaining sufficient physical activity. Med Sci Sports Exerc.2012; 44(3): 466-473. Epub 2011/08/17. doi: 10.1249/MSS.0b013e318230e441 PMID: 21844823
30. Sagelv EH, Ekelund U, Pedersen S, Brage S, Hansen BH, Johansson J, Grimsgaard S, Nordstrom A, Horsch A, Hopstock LA, Morseth B. Physical activity levels in adults and elderly from triaxial and uniaxial accelerometry. The Tromsø Study. PloS ONE.2019; 14(12): e0225670. https://doi.org/10.1371/journal.pone.0225670.
31. Hansen BH, Kolle E, Johannessen JS, Dalene KE, Ekelund U, Anderson SA. Monitoring population levels of physical activity and sedentary time in Norway across the lifespan. Scand J Med Sci Sports.2019; 29(1): 105–112. https://doi.org/10.1111/sms.13314 PMID: 30276928
32. Troiano RP, Berrigan D, Dodd KW, Masse LC. Tillert, T., Mcdowell, M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc.2008; 40(1): 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3 PMID: 18091006.
33. Berkemeyer K, Wijndarle K, White T, Cooper AJM, Luben R, Westgate K, Griffin SJ, Khaw KT, Wareham NJ, Brage S. The descriptive epidemiology of accelerometer-measured physical activity in older adults. Int J Behav Nutr Phys Act.2016; 13, p. 2. https://doi.org/10.1186/s12966-015-0316-z
34. Doherty A, Jackson D, Hammerla N, PloÈtz T, Olivier P, Granat MH, Whit. T, Hees VT, Trene MI, Owem G, Preece SJ, Gillions R, Peakman S, Brage S, Wareham NJ. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. PLoS ONE.2017; 12(2): e0169649. doi:10.1371/ journal.pone.0169649.
35. Schaller N, Seiter H, Himmerich S, Karg G, Gedrich K, Wolfram G, Linseisen J. Estimated physical activity in Bavaria, Germany, and its implications for obesity risk: results from the BVS-II study. Int J Behav Nutr Phys Act.2005; 2(1): 1-6. DOI:10.1186/1479-5868-2-6.
36. Jaeschke L, Steinbrecher A, Boeing H, Gastell S, Ahrens W, Berger K, Brenner H, Ebert N, Fischer B, Greiser KH, Hoffmann W, Jöckel KH, Kaaks R, Keil T, Kemmling Y, Kluttig A, Krist L, Leitzmann M, Lieb W, Linseisen J, Löffler M, Michels KB, Obi N, Peters A, Schipf S, Schmidt B, Zinkhan M, Pischon T. Factors associated with habitual time spent in different

physical activity intensities using multiday accelerometry. Scientific Reports.2020; 10, 774 | https://doi.org/10.1038/s41598-020-57648-w
37. Levine JA, Lanningham- Foster LM, McCrady SK, Krizan AC, Olson LR, Kane PH, Jensen MD, Clark MM. Inter-individual variation in posture allocation: possible role in human obesity. Scence.2005; 307: 584-586.
38. Bann D, Hire D, Manini T, Cooper R, Botoseneanu A, McDermott MM, Pahor M, Glynn NW, Fielding R, King AC, Church T, Ambrosius WT, Gill T. Light Intensity Physical Activity and Sedentary Behavior in Relation to Body Mass Index and Grip Strength in Older Adults: Cross-Sectional Findings from the Lifestyle Interventions and Independence for Elders (LIFE) Study. PLoS ONE.2015; 10(2): e0116058. doi:10.1371/journal.pone.0116058.
39. Thorp AA, Helly GN, Winkle E, Clark B. Prolonged sedentary time and physical activity in workplace and non-work contexts: a cross-sectional study of office, customer service and call centre employees. Int J Behav Nutr Phys Act.2012; 9, 128. https://doi.org/10.1186/1479-5868-9-128 PMID: 23101767.
40. Kantomaa MT, Tikanmaki M, Kankaanpaa A, Vaarasmaki M, Sipola-Leppanen M, Ekelund U, Hakonen H, Jarvelin MR, Kajantie E, Tammelin TH. Accelerometer-Measured Physical Activity and Sedentary Time Differ According to Education Level in Young Adults. PLoS One.2016; 11(7): e0158902. https://doi.org/10.1371/ journal.pone.0158902 PMID: 27403958
41. Hansen BH, Ommundsen Y, Holme I, Kolle E, Anderssson SA. Correlates of objectively measured physical activity in adults and older people: a cross-sectional study of population-based sample of adults and older people living in Norway. Int J Public Health.2014; 59(2): 221–230. https://doi.org/10.1007/s00038-013-0472-3 PMID: 23619723
42. Holtermann A, Krause N, Beek AJ, Straker L. The physical activity paradox: six reasons why occupational physical activity (OPA) does not confer the cardiovascular health benefits that leisure time physical activity does. Br J Sports Med.2018; 52(3): 149–150. https://doi.org/10.1136/bjsports-2017-097965 PMID: 28798040
43. Bernaards CM, Twisk JW, Van Mechelen W, Snel J, Kemper HCA. Longitudinal study on smoking in relationship to fitness and heart rate response. Med Sci Sports Exerc.2003; 35: 793–800.
44. Spring B, Moller AC, Coons M J. Multiple health behaviours: overview and implications. J Public Health (Oxf).2012; 34: 3–10
45. St-Onge MP, Grandner MA, Brown D, Corny MB, Coons M, Bhatt DP. Sleep duration and quality: impact on lifestyle behaviors and cardiometabolic health: a scientific statement from the American Heart Association. Circulation.2016; 134(18): 367–386. doi:10.1161/CIR.0000000 000000444
46. Medina C, Barquera S, Janssen I. Validity and reliability of the International Physical Activity Questionnaire among adults in Mexico. Rev Panam Salud Publica.2013; 34(1): 21–8.

Table 1. Descriptive characteristics of the participants.

Variables	Subcategory	Total sample	Men	Women
		(N=97)	**(n=30, 30.9)**	**(n=67, 69.1)**
Age (years)	Mean (±SD)	38.9±8.9	44.3±6.9	36.4±8.7
Age group (n, %) *	18-29	17 (17.5)	0 (0.0)	17 (25.4)
	30-45	55 (56.7)	16 (53.3)	39 (58.2)
	>45	25 (25.8)	14 (46.7)	11 (16.4)
Marital status (n, %) *	Single	25 (25.8)	2 (6.7)	23 (34.3)
	Married	72 (74.2)	28 (93.3)	44 (65.7)
BMI (Kg/m2)	Mean (±SD)	27.6±4.7	28.8±3.3	27.1±5.2
BMI Category (n, %) *	Normal Weight	30 (30.9)	4 (13.3)	26 (38.8)
	Overweight/Obese	67 (69.1)	26 (86.7)	41 (61.2)
Educational level (n, %)	< Secondary School	12 (12.4)	5 (16.7)	7 (10.4)
	Secondary School	17 (17.5)	8 (26.7)	9 (13.4)
	> Secondary School	68 (70.1)	17 (56.7)	51 (76.1)
Smoking (n, %)	Yes	41 (42.3)	16 (53.3)	25 (37.3)
	No	56 (57.7)	14 (46.7)	42 (62.7)

*Significant difference exists between men and women at p<0.05
BMI: Body Mass Index
N: number of subjects

Table 2. Descriptive intensity of activity and step counts, by sex, age, marital status, BMI, education and smoking; Mean (SD).

		intensity of activity						
Variables/Subcategory	**N**	**Sedentary**	**Light**	**Moderate**	**vigorous**	**MVPA with bouts**	**MVPA without bouts**	**Steps per day**
		Minutes per day (mean (SD))						**Average**
Sex								
Total	97	644.2 (90.7)	196.0 (40.8)	41.0 (19.7)	1.0 (1.3)	7.3 (10.9)	42.0 (20.1)	7502.2 (2513.2)
Women	67	641.0 (89.7) [a]	193.8 (44.1) [a]	35.5 (14.5) [a]	0.8 (1.1) [a]	5.5 (8.1) [a]	36.3 ± (14.9) [a]	6904.1 (2111.3) [a]
Men	30	651.3 (94.0) [a]	200.7 (32.5) [a]	53.2 (24.0) [b]	1.5 (1.7) [b]	11.5 (14.9) [b]	54.7 (24.3) [b]	8838.0 (2844.4) [b]
Age (years)								
18-29	17	670.6 (136.9) [a]	175.6 (41.4) [a]	36.3 (12.0) [a]	0.9 (1.0) [a,b]	6.1 (10.9) [a]	37.2 (12.3) [a]	6966.9 (1791.9) [a]
30-45	55	629.7 (68.6) [a]	198.3 (40.6) [b]	39.5 (15.5) [a]	0.9 (1.1) [a]	6.1 (8.1) [a]	40.3 (15.8) [a]	7449.8 (2201.4) [a]
>45	25	658.2 (93.0) [a]	204.7 (37.9) [b]	47.5 (29.0) [a]	1.5 (1.8) [b]	9.7 (15.7) [a]	49.0 (29.6) [a]	7981.4 (3430.9) [a]
Marital status								
Single	25	659.8 (116.4) [a]	186.5 (44.1) [a]	41.0 (14.6) [a]	1.1 (1.4) [a]	8.7 (10.6) [a]	42.1 (15.1) [a]	7888.7 (2424.8) [a]
Married	72	638.8 (80.2) [a]	199.3 (39.4) [a]	41.0 (21.2) [a]	1.0 (1.3) [a]	6.9 (11.1) [a]	41.9 (21.7) [a]	7368.0 (2545.9) [a]
BMI (Kg/m2)								
Normal Weight	30	646.1 (111.2) [a]	199.7 (51.7) [a]	45.3 (16.0) [a]	1.0 (1.2) [a]	8.3 (9.5) [a]	46.3 (16.1) [a]	7960.7 (2468.2) [a]
Overweight/Obese	67	643.4 (80.8) [a]	194.3 (35.2) [a]	39.1 (20.9) [a]	1.0 (1.4) [a]	6.9 (11.6) [a]	40.1 (21.5) [a]	7296.9 (2524.3) [a]
Educational level								
< Secondary School	12	599.5 (71.5) [a]	205.0 (35.0) [a]	53.4 (18.7) [a]	1.4 (1.4) [a]	11.3 (9.6) [a]	54.7 (18.9) [a]	9277.0 (3280.0) [a]
Secondary School	17	644.7 (88.0) [a]	191.7 (34.4) [a]	39.0 (18.3) [a,b]	1.2 (1.5) [a]	6.3 (7.3) [a]	40.1 (19.0) [a,b]	7032.1 (2695.0) [b]
> Secondary School	68	652.0 (93.1) [a]	195.5 (43.4) [a]	39.3 (19.6) [b]	1.0 (1.3) [a]	6.9 (11.8) [a]	40.2 (20.0) [b]	73.6.5 (2211.4) [b]
Smoking								
Yes	41	633.0 (75.0) [a]	205.0 (35.0) [a]	41.0 (21.0) [a]	1.3 (1.6) [a]	6.0 (7.2) [a]	42.3 (21.6) [a]	7499.9 (2563.6) [a]
No	56	652.4 (100.5) [a]	195.7 (38.6) [a]	40.9 (18.8) [a]	1.0 (1.1) [a]	8.3 (13.0) [a]	41.8 (19.1) [a]	73.6.5 (2499.0) [a]

* Significant difference< 0.05
SD: standard deviation
BMI: Body Mass Index
N: number of subjects

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Figure 1. The proportion of women, men, total and by age groups, fulfilling the WHO's recommendations for physical activity of 150 minutes of MVPA per week, in both non-bouted and bouted MVPA.

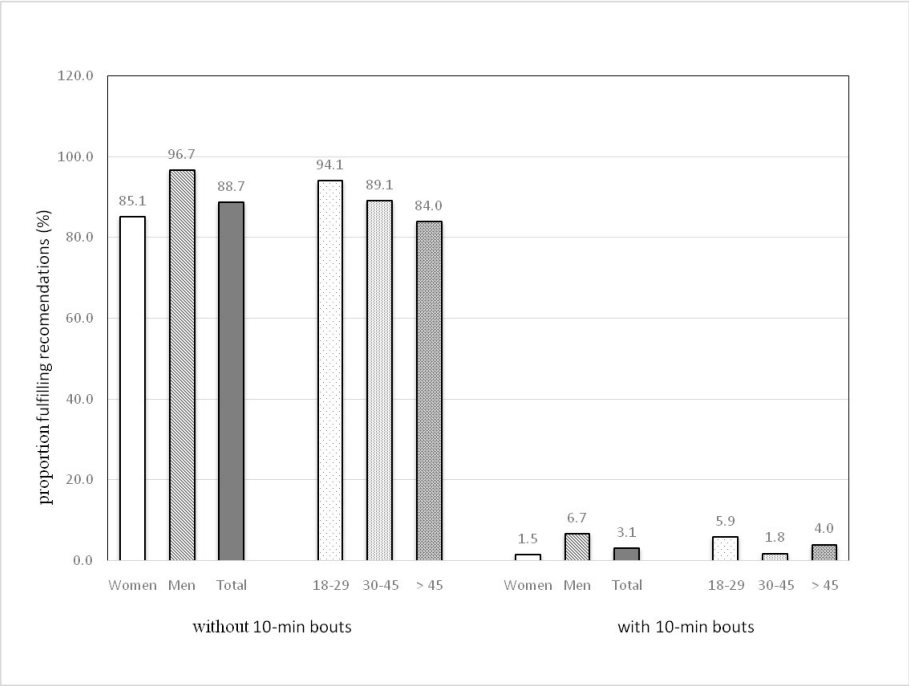
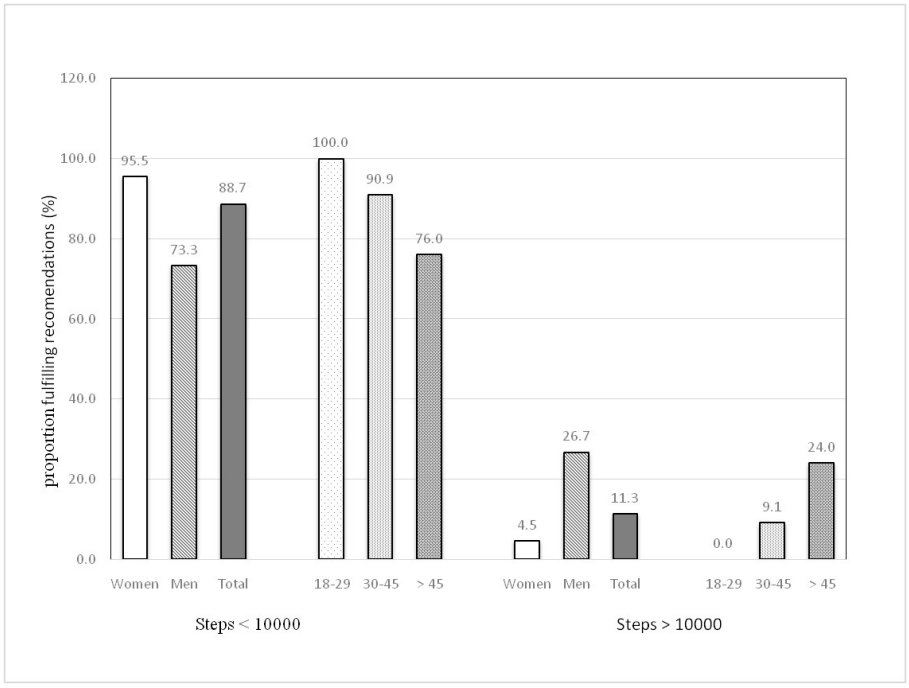


Figure 2. The proportion of women, men, total and by age groups, fulfilling the WHO's recommendations for step counts of 10000 steps per day



ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Potential Biases of the Transmission Risks of COVID-19 estimated by Contact Tracing Surveys in Japan

Tsubasa Ito[(1)], Takahiro Otani[(2)], Tatsuhiko Anzai[(3)], Takashi Okumura[(4)], Kunihiko Takahashi[(3)]

(1) Faculty of Public Policy, Graduate School of Public Policy, Hokkaido University, Hokkaido, Japan
(2) Department of Public Health, Graduate School of Medical Sciences, Nagoya City University, Aichi, Japan
(3) Department of Biostatistics, M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan
(4) Health Administration Center, Kitami Institute of Technology, Hokkaido, Japan

CORRESPONDING AUTHOR: Tsubasa Ito, Faculty of Public Policy, Graduate School of Public Policy, Hokkaido University, Hokkaido, Japan. E-mail: tito@econ.hokudai.ac.jp, ORCID 0000-0001-8021-9907

SUMMARY

Introduction: Contact tracing surveys are being conducted to identify and isolate close contacts of an identified patient to reduce the spread of coronavirus disease (COVID-19). However, the estimates of risk indexes based on information obtained from the surveys and normally used in practice can have biases comparing with true magnitude of risks of infection and spread.
Method: We evaluated whether the estimates of the risk indexes obtained from information of the active epidemiological surveillance, contact tracing surveys in Japan, are suitable for quantitative assessment of the risk factors of COVID-19, using pseudo data via a simulation study. We discussed two types of risks considered in the issue of infectious disease, the probability of infection and that of spreading, and the estimates of these risks.
Results and Discussion: A naive method to estimate the risks of infection and spreading of COVID-19 is to calculate the ratio of infected patients to close contacts and the ratio of patients who infected others to all the confirmed patients, respectively. However, these estimates could possibly have significant biases and result in being ineffective for both the exploration and the quantitative assessment of the risk factors in the following ordinary cases: a person contacts closely with many confirmed patients, or a confirmed patient contact closely with many people. Then, some steps are needed to reduce such possible biases for the estimation the risks of both the infection and spreading of COVID-19.

Keywords: spatio-temporal epidemiology; COVID-19; active epidemiological surveillance; evidence-based policy-making; infection risk; spreading risk; simulation study.

INTRODUCTION

The spread of the coronavirus disease (COVID-19) still can threaten the global health, and the recent (1 April, 2023) World Health Organization (WHO) COVID-19 dashboard has reported approximately 763 million confirmed cases and 7 million deaths globally [1].

Many countries use contact tracing as one of the most powerful public health interventions. The common purpose of these tracings is to identify and isolate individuals who may have been infected due to close contact with an identified patient, to prevent the infectious disease from being transmitted further [2, 3]. In Japan, a bidirectional contact tracing called "active epidemiological surveillance" is being conducted [4]. The survey investigates not only the close contacts of a confirmed patient but also other patients who possibly infected the patient, i.e. sources of infection, and then successive contact tracings are iterated to identify additional patients and close contacts related to the patient [5].

Many researchers have investigated COVID-19 using information from contact tracings, which have enabled us to identify risk factors for infection and spread. For example, it is well-known that being in a closed and poorly ventilated environment causes higher secondary transmissions of COVID-19 than

DOI: 10.54103/2282-0930/20757
Accepted: 16th June 2023

being in an open and well-ventilated environment [6, 7]. Furthermore, in Japan, the data from active epidemiological surveillance have been used to implement COVID-19 measures, and some new phrases have been coined, such as "stay away from the three Cs (closed spaces, crowded places, close-contact settings) " and "five situations that increase the risk of infection (social gatherings involving alcohol consumption, big feasts in large groups, conversation without a mask, living together in a small limited space, and moving to different locations)" [8, 9], which has been considered to contribute significantly to restrain the pandemic.

The risks of infection and spread are generally defined as the probability that a closed contact becomes infected and the probability that a patient infects the other, respectively. Then, in the standard epidemiological investigations, these risks are typically estimated by the ratio of the number of confirmed positive patients to the number of identified close contacts and the ratio of the number of patients who are identified to infect others to the number of confirmed patients as discussed in the next section. Although these simple estimates are practical, it is likely that they have biases and could not reflect the true magnitude of risks of infection and spread. For example, if a confirmed patient with some risk factor (e.g., without a mask) becomes close contacts with someone and they develop COVID-19, the confirmed patient will be seen as the spreader even if they actually did not infect others. Then, if one person becomes contact with a number of patients with or without symptoms, which is likely enough, that risk factor would be overestimated.

Thus, this study aimed to investigate how these estimates of the risk indexes based on information obtained by the active epidemiological surveillance could have biases comparing with the true magnitude of the risk factors for COVID-19 via simulation studies under some situations reflecting how people come into contact with each other. Assessment of potential biases of risk estimates helps improve the estimations of risk of infection and spread, which tells how we should act to prevent COVID-19 from spreading, recognizing the limitations of the risk estimates technically feasible.

METHODS

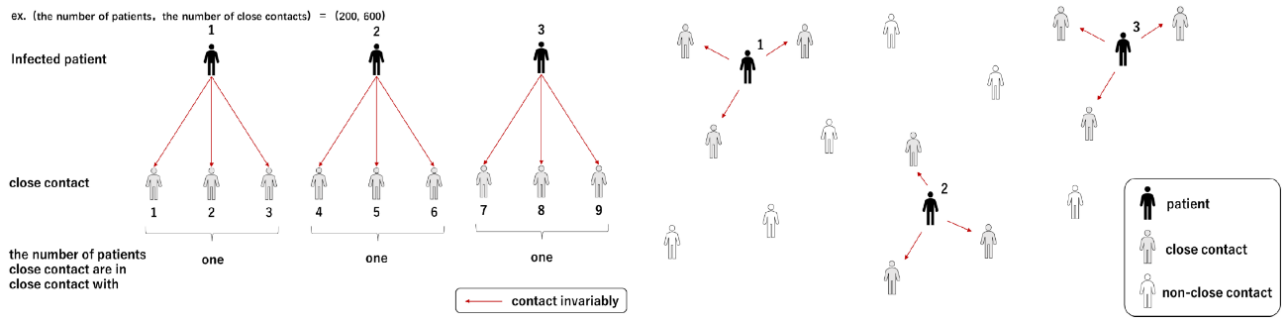
Simulation study

In this section, by using simulation studies we investigated the performances of the estimates of risks for quantitative risk assessment under two scenarios.

In the first scenario, to generate pseudo data, we followed the form of the active epidemiological surveillance in Japan. This investigation targets patients, suspected disease carriers, and their close contacts [10]. From this investigation, we can extract spatiotemporal information about the behaviors of patients (when, where, with whom, what and how they did). Then, we supposed several activities, such as watching sports in a stadium or seeing a live concert in a venue, and we set the population size, not including the infected individuals who were unaware of their condition, to 10,000. We assumed that the prevalence rate of COVID-19 in Japan is approximately 2.0% based on the estimate of the average detection rate of COVID-19 [11]. Thus, the number of patients was set to 200. We considered the case in which an individual participates in an event alone, with some or many people. One patient was in close contact with at least 1 person or 3 or 5 people, thus, there are three values for the number of close contacts: 200, 600, or 1,000 close contacts. We also considered patterns concerning how large the flow of people is. Thus, three patterns were established for the number of patients a close contact comes in close contact with on average: 1, 1.2, and 1.5. These settings can be interpreted as follows. When the number of patients per close contact on average was 1, the patients can be considered to be distributed uniformly and the infections occur globally, as shown in Figure 1.

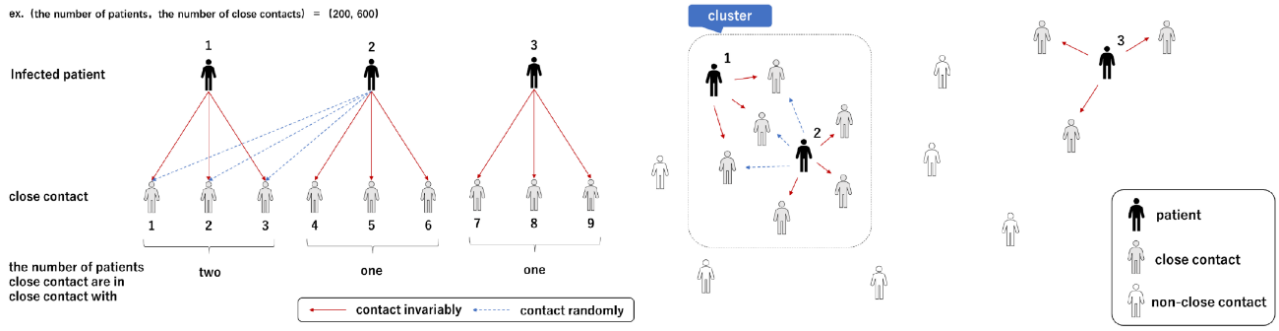
In contrast, when the average number of patients per close contact was 1.2, we randomly selected a patient with whom a close contact has come in close contact. For example, as a result of random selection, if close contact 1 came in close contact with patient 2 in addition to patient 1, we assume that close contacts 2 and 3 also came in close contact with patient 2. In this case, a number of patients were concentrated in

Figure 1: Scenario in which the infection occurs globally (the number of patients per close contact is fixed) (left panel), and illustration for this scenario (right panel).



ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Figure 2: Scenario in which the infection occurs locally (left panel), and illustration for this scenario (right panel).



one portion and the infections occurred locally; this formed a cluster, as shown in Figure 2.

Moreover, as the average number of patients per close contact increased to 1.5, more close contacts were more likely to come in close contact with a common patient, and the size of the cluster became larger.

In this study, we considered scenarios in which individuals are exposed if they did not wear a mask. For simplicity, the probability of patients who do not wear a mask and the probability that a close contact did not wear a mask were set to 0.5. When patients did not wear a mask, the probability that their close contacts who also did not wear a mask would be infected was set to 4/9, while the probability that their close contacts with masks would be infected was set to 2/9. Contrarily, when patients wear a mask, the probability that their close contacts without masks would be infected was set to 2/9, while the probability that their close contacts with masks would be infected was set to 1/9. Therefore, the true RR of the infection and spreading risks was 2.0, and the true OR of the infection and spreading risks was 2.5. If individuals were in close contact with many patients and they were infected, it is generally difficult to infer who infected the individual. Therefore, in such cases, we considered that the individual was infected from all patients with whom they were in close contact.

In the second scenario, the simulation setting is almost same with that of the first scenario. In this time, however, the average number of close contacts per patient were not fixed, and we considered six patterns for it, that is, 1.5, 2.0, 3.5, 4.5, 5,0 and 6.0, and we randomly selected people who became in close contact with each patient such that all the patients were in close contact with at least one person. Then, the interpretation of this setting is similar to that of the case in which the number of patients close contacts are in close contact with on average is 1 in the first scenario. In this time, however, the number of close contacts a patient is in close contact with varies depending on the patient, and then this setting is closer to the realistic situation.

Statistical Analysis

We estimated the risks of both the infection and spreading of COVID-19 associated with human behaviors, which were whether they wore masks in a crowd of people in our simulation setting.

In each scenario, we estimated the infection risk and the spreading risk respectively by

$$\text{infection risk} \approx \frac{\text{the number of confirmed positive patients}}{\text{the number of identified close contacts}}$$

and

$$\text{spreading risk} \approx \frac{\text{the number of patients who are identified to infect others}}{\text{the number of confirmed patients}}$$

for both the exposed and unexposed groups. The exposed group comprises patients or close contacts who practiced risky behaviors that increased the likelihood of developing an infection, such as not wearing masks. Then, we calculated the relative risk (RR) or odds ratio (OR) as the ratio of the risk of the two groups for both the infection risk and the spreading risk. Moreover, we constructed 95% confidence intervals (CI) of each RR and OR. Then, based on 2,000 simulation runs, we computed the averaged values of the RR and OR of the infection and spreading risks, the standard deviations of the RR and OR, and the averaged values of coverage probabilities (CPs) of 95% confidence intervals (CI) of the infection and spreading risks.

Results

The results of the first scenario of the simulation study are reported in Tables 1 and 2. The RR and OR of the infection and spreading risks performed the best when both the number of patients and the number of close contacts were 200 and the average number of patients per close contact was 1. [RR=2.08, CP of its 95% CI is 95.6%, OR=2.67 and CP of its 95% CI is 95.6% for infection risk, and RR=2.11, CP of its 95% CI is 94.9%, OR=2.71 and CP of its 95% CI is 94.7% for spreading risk]

On the other hand, when both the number of patients and the number of close contacts were 200 and the average number of patients per close contact increased to 1.5, for example, RR=1.94, the CP of its 95% CI is 93.5%, OR=2.71 and the CP of its 95% CI is 95.1% for infection risk and, for RR=1.37, the CP of its 95% CI is 27.6%, OR=1.85 and the CP of its 95% CI is 78.6% for spreading risk.

Moreover, when the number of close contacts increased to 1,000 for example and the average number of patients per close contact was 1, for RR=2.02, the CP of its 95% CI is 94.6%, OR=2.53 and the CP of its 95% CI is 94.6% for infection risk and, for RR=1.46, the CP of its 95% CI is 9.8%, OR=4.86 and the CP of its 95% CI is 64.5% for spreading risk.

The results of the second scenario are reported in Tables 3 and 4. The values of RR of infection risk and its CP of 95% CI corresponding to the average number of close contacts per patient, 1.5, 2.0, 3.5, 4.5, 5,0 and 6.0, are (RR, CP)=(2.07, 94.8%), (2.04, 94.8%), (2.01, 94.9%), (2.00, 96.1%), (2.00, 95.8%) and (2.00, 93.7%), respectively. The values of OR of infection risk and its CP of 95% CI corresponding to the average number of close contacts per patient, 1.5, 2.0, 3.5, 4.5, 5,0 and 6.0, are (OR, CP)=(2.64, 95.0%), (2.59, 94.4%), (2.55, 94.9%), (2.54, 95.6%), (2.54, 95.3%) and (2.54, 94.0%), respectively. On the other hand, the values of RR of spreading risk and its CP of 95% CI corresponding to the average number of close contacts per patient, 1.5, 2.0, 3.5, 4.5, 5,0 and 6.0, are (RR, CP)=(1.90, 92.5%), (1.78, 86.1%), (1.54, 39.7%), (1.42, 9.5%), (1.37, 2.9%) and (1.30, 0.0%), respectively. The values of OR of spreading risk and its CP of 95% CI corresponding to the average number of close contacts per patient, 1.5, 2.0, 3.5, 4.5, 5,0 and 6.0, are (RR, CP)=(2.64, 95.8%), (2.70, 95.8%), (3.05, 92.4%), (3.34, 90.6%), (3.50, 90.0%) and (4.03, 87.2%), respectively. Then, we can see that when the average number of close contacts per patient is large, the CI of RR for spreading risk did not work at all.

Table 1: The averaged values of the relative risk (RR), odds ratio (OR) and coverage probabilities (CP) of their 95% confidence intervals of the infection risk, the standard deviation in parentheses and α is the number of patients with whom a close contact becomes close contact on average.

		RR	CP for RR	OR	CP for OR
	α = 1	2.08 (0.77)	95.6%	2.67 (0.99)	95.6%
(200,200)	α = 1.2	2.05 (0.74)	95.2%	2.73 (0.99)	95.6%
	α = 1.5	1.94 (0.67)	93.4%	2.71 (0.96)	95.1%
	α = 1	2.03 (0.57)	95.4%	2.56 (0.73)	95.0%
(200,600)	α = 1.2	1.97 (0.53)	94.8%	2.56 (0.70)	95.4%
	α = 1.5	1.89 (0.51)	91.8%	2.59 (0.70)	95.0%
	α = 1	2.02 (0.49)	94.6%	2.53 (0.63)	94.6%
(200,1000)	α = 1.2	1.96 (0.47)	94.1%	2.54 (0.62)	95.0%
	α = 1.5	1.87 (0.47)	88.1%	2.56 (0.61)	94.8%

Table 2: The averaged values of the relative risk (RR), odds ratio (OR) and coverage probabilities of their 95% confidence intervals (CP) of the spreading risk, the standard deviation in parentheses and α is the number of patients with whom a close contact becomes close contact on average.

		RR	CP for RR	OR	CP for OR
	α = 1	2.11 (0.81)	94.9%	2.71 (1.03)	94.7%
(200,200)	α = 1.2	1.64 (0.71)	78.9%	2.14 (0.87)	89.9%
	α = 1.5	1.37 (0.82)	27.6%	1.85 (0.92)	78.6%
	α = 1	1.69 (0.62)	70.9%	3.46 (1.21)	85.1%
(200,600)	α = 1.2	1.35 (0.81)	2.6%	2.98 (1.08)	94.6%
	α = 1.5	1.15 (0.92)	0.0%	2.99 (1.30)	95.4%
	α = 1	1.46 (0.75)	9.8%	4.86 (1.76)	64.5%
(200,1000)	α = 1.2	1.19 (0.90)	0.0%	4.95 (2.03)	88.0%
	α = 1.5	1.06 (0.97)	0.0%	– (–)	–

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Table 3: ncc is mean of the number of close contacts, "mean" is mean of the number of close contacts per patient and "sd" is its standard deviation. Averaged values of the relative risk (RR), odds ratio (OR), the standard deviations in parentheses and coverage probabilities (CP) of their 95% confidence intervals of the infection risk.

ncc	(mean, sd)	RR	CP for RR	OR	CP for OR
297.6	(1.5, 0.70)	2.07 (0.71)	94.8%	2.64 (0.92)	95.0%
394.2	(2.0, 1.05)	2.04 (0.65)	94.8%	2.59 (0.83)	94.4%
678.1	(3.5, 1.58)	2.01 (0.54)	94.9%	2.55 (0.83)	94.9%
863.2	(4.5, 1.87)	2.00 (0.51)	96.1%	2.54 (0.65)	95.6%
953.0	(5.0, 2.00)	2.00 (0.49)	95.8%	2.54 (0.63)	95.3%
1131.9	(6.0, 2.24)	2.00 (0.47)	93.7%	2.54 (0.61)	94.0%

Table 4: ncc is mean of the number of close contacts, "mean" is mean of the number of close contacts per patient and "sd" is its standard deviation. Averaged values of the relative risk (RR), odds ratio (OR), the standard deviations in parentheses and coverage probabilities (CP) of their 95% confidence intervals of the spreading risk.

ncc	(mean, sd)	RR	CP for RR	OR	CP for OR
297.6	(1.5, 0.70)	1.90 (0.67)	92.5%	2.64 (0.94)	95.8%
394.2	(2.0, 1.05)	1.78 (0.62)	86.1%	2.70 (0.92)	95.8%
678.1	(3.5, 1.58)	1.54 (0.71)	39.7%	3.05 (1.07)	92.4%
863.2	(4.5, 1.87)	1.42 (0.77)	9.5%	3.34 (1.20)	90.6%
953.0	(5.0, 2.00)	1.37 (0.80)	2.9%	3.50 (1.28)	90.0%
1131.9	(6.0, 2.24)	1.30 (0.84)	0.0%	4.03 (1.56)	87.2%

DISCUSSION

This study evaluated the estimates of the risks of infection and spreading of COVID-19 obtained using information from an active epidemiological surveillance.

From the simulation results, we can deduce the following findings. At first, we fix the number of patients and that of close contacts, and varies the number of the patients per close contact. (ex. We see the case where the number of patients is 200 and the number of close contacts is 600, and varies α from 1 to 1.5 in Table 1.) Then, we see that the performances of the both RR and OR of the infection risk were stable for most cases. However, in the case where the number of patients is 200, the number of close contacts is 1000 and the number of the patients per close contact on average is 1.5, RR has a little downward bias and the CP of 95% confidence interval is much smaller than the nominal confidence level. This downward bias might come from the fact that when a person is in close contact with many patients, they are more likely to be infected from one of the patients even if the patient wore a mask, which increases the infection risk of unexposed group. For the spreading risk, their RR had downward biases which became large as the average number of patients per close contact. Concerning the OR, slightly downward biases arose when the number of close contacts was 200 and large upward biases arose when the number of close contacts was 600 or 1000. Consequently, their Cis did not achieve the nominal confidence level at all. These different directions of large biases might be explained by the same reason. An individual comes in close contact with many patients, the patients are more likely to be classified in the group of patients who infected others no matter whether they actually infected others. This increases the spreading risk for the unexposed group of patients, which causes underestimation of the RR of spreading risk. Simultaneously, the estimates of the risk of spreading risk for the exposed group were close to 1, which causes the overestimation of OR.

Next, we fix the number of the patients per close contact, and varies the number of close contacts. (ex. We see the case where α is 1, and varies the number of close contacts from 200 to 1000 in Table 1.) Then, the performances of the RR and OR of the infection risk were considerably stable, but for the spreading risk, large downward biases occurred on the RR and large upward biases in the OR were observed as the number of close contacts increased. These findings are consistent with the results of the second scenario. Both the RR and OR for the infection risk were accurate enough, because almost all close contacts came in close contact with one patient (Table 3), and the RR and OR for the spreading risk were not accurate enough as the average number of close contacts per patient increased (Table 4). These downward biases

might be because when patients come in contact with many people, they are more likely to infect one of their close contacts even if they wore a mask, and this results in an increase in the spreading risk for the unexposed group of patients.

All these poor performances of RR and OR might be reasonable because these estimates are only based on the number of close contacts. However, we might be able to evaluate the direction of biases in the estimates of infection or spreading risks by investigating the average number of patients per close contact or the number of close contacts per patient, which might be feasible by using the methods which evaluate the infection risk using the locational information of mobile phones proposed by [12] and [13] for example.

Practically, there are other factors that can cause biases in risk estimators. For example, we cannot exactly trace all close contacts using the active epidemiological surveillance [14], which causes a selection bias. This is because the travel recall of patients may be inaccurate or indistinct, some patients may be uncooperative during the investigation, patients with subjective symptoms fail to present at medical institutions, and patients with no or mild symptoms do not perceive as having the infection. Therefore, the reported number of close contacts is usually smaller than the true number. The reported number of patients is also smaller than the true number, because it has been shown that a significant proportion of patients are asymptomatic [15]. These factors cause bias in the estimation of the infection risk.

Lastly, the prevalence rate in the group of close contacts tends to be high; therefore, the OR computed from case-controlled studies using information from an active epidemiological surveillance might be higher than the actual RR.

Limitations

In this study, we could not consider all situations that could actually occur in the simulation setting and could only consider certain simple situations. In our simulation setting, the probabilities that two individuals will come in close contact with each other are the same for all patients and close contacts, though these probabilities depend on many confounding factors. Moreover, in our simulation setting, all patients and their close contacts were completely matched, though such exact tracing is impossible. In these more complicated situations, it can be expected that the RR and OR will not be as accurate.

CONCLUSION

In the future, more complicated simulation studies should be conducted, specifically, those that consider the movement of people. The simple simulations performed in this study, however, suggested some issues of the estimates of the risks of infection and spread of COVID-19. There were a few possible biases in the estimates of infection risks. The RR and OR based on these estimates are useful in the identification and quantitative assessment of the risk factors for infection associated with human behaviors. However, there were several possible biases in the estimates of spreading risks. Thus, we cannot use the RR and OR based on the estimates of the spreading risk for that purpose. Moreover, for more complicated situations, we might not be able to use the estimates of both infection and spreading risks if we simply calculate the estimates of the risks using the information from an active epidemiological surveillance. In this case, we should collect more high-quality information such as the number of times of contacts in contact tracing surveys.

FUNDING

This research was supported by the Japan Agency for Medical Research and Development (AMED) under Grant Number JP20he0622042.

REFERENCES

1. World Health Organization. Who coronavirus disease (covid-19) dashboard, 2023.
2. J Cohen and Kupferschmidt K. Countries test tactics in 'war' against covid-19. *Science*, 367:1287–1288, 2020.
3. Emeline Han MSc, Melisa Mei Jin Tan MSc, Eva Turk PhD, Devi Sridhar PhD, Gabriel M Leung MD Kenji Shibuya DrPH, Nima Asgari MD Juhwan Oh PhD, Alberto L Garcia-Basteiro PhD, Johanna Hanefeld PhD, Alex R Cook PhD, Li Yang Hsu MBBS, Yik Ying Teo PhD, David Heymann DTM&H, Helen Clark MA, Martin McKee DSc, and Helena Legido-Quigley PhD. Lessons learnt from easing covid-19 restrictions: an analysis of countries and regions in asia pacific and europe. *Lancet, Health Policy*, 396:1525–1534, 2020.
4. Ministry of Health, Labor, and Welfare, Japan. Preventing the spread of covid-19 by targeting disease clusters, 2020.
5. Sadamori Kojaku, Laurent Hebert-Dufresne, Enys Mones, Sune Lehmann, and Yong-Yeol Ahn. The effectiveness of backward contact tracing in networks. *nature physics*, 2021.
6. Hiroshi Nishiura, Hitoshi Oshitani, Tetsuro Kobayashi, Tomoya Saito, Tomimasa Sunagawa, Tamano Matsui, Takaji Wakita, MHLW COVID-19 Response Team, and Motoi Suzuki. Closed environments facilitate secondary transmission of coronavirus disease 2019 (covid-19). *medRxiv*, 32(12), 2020.
7. Valentyn Stadnytskyi, Christina E. Bax, Adriaan Bax, and Philip Anfinrud. The airborne lifetime of small speech droplets and their potential importance in

sars-cov-2 transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (22), 2020.
8. Ministry of Health, Labour and Welfare, Japan. Avoid the "three cs", 2021.
9. Ministry of Health, Labour and Welfare, Japan. "5 situations" that increase the risk of infection, 2021.
10. National Institute of Infectious Diseases. Guidelines for active epidemiological investigation in patients with novel coronavirus infection (tentative version), 2020.
11. Sebastian Vollmer. Average detection rate of sars-cov-2 infections has improved since our last estimates but is still as low as nine percent on march 30th, 2020.
12. Junko Ami, Kunihiro Ishii, Yoshihide Sekimoto, Hiroshi Masui, Ikki Ohmukai, Yasunori Yamamoto and Takashi Okumura. Computation of infection risk via confidential locational entries: A precedent approach for contact tracing with privacy protection. *IEEE Access*, 9: 87420 – 87433, 2021
13. Ikki Ohmukai, Yasunori Yamamoto, Maori Ito and Takashi Okumura. Tracing patients' PLOD with mobile phones: Mitigation of epidemic risks through patients' locational open data. IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE-2020), June 2020. arXiv:2003.06119.
14. Matt J Keeling, Deirdre Hollingsworth, and Jonathan M Read. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (covid-19). *Journal of Epidemiology and Community Health*, 74:861–866, 2020.
15. Daniel P Oran and Eric J Topol. The proportion of sars-cov-2 infections that are asymptomatic : A systematic review. *Annals of internal medicine*, 2021.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Choosing a medical specialty course in Italy: explorative analysis of factors related to the choice

Paolo Trerotoli[(1)], Massimo Minerva[(2)], Nicola Bartolomeo[(3)], Massimo Giotta[(4)], Ilaria Dargenio[(4)], Giammaria Liuzzi[(5)], Carlo Signorelli[(6)]

(1) Associate Professor, Medical Statistic, Department of Interdisciplinary Medicine, University of Bari Aldo Moro, Italy
(2) School of Medicine, Vita-Salute San Raffaele University, Milan, Italy; University of Parma, Parma, Italy
(3) Department of Interdisciplinary Medicine, University of Bari Aldo Moro, Italy
(4) School of Specialization in Medical Statistics and Biometry, Department of Interdisciplinary Medicine, University of Bari Aldo Moro, Italy
(5) Laboratory Medicine, ASST Rhodense, Milan, Italy
(6) School of Medicine, Vita-Salute San Raffaele University, Milan, Italy; University of Parma, Parma, Italy

CORRESPONDING AUTHOR: Paolo Trerotoli, Department of Interdisciplinary Medicine, Piazza Giulio Cesare 11 70124 Bari, University of Bari Aldo Moro, Italy. E-mail: paolo.trerotoli@uniba.it, ORCID 0000-0001-7102-9108

SUMMARY

Background: In order to be able to access a course of medical or surgical specialization, in addition to the degree and the qualification to the profession it is necessary to perform an entrance test. In this study we wanted to analyze the possible factors that determined the choice of a given graduate school and the place where to attend it in the year 2021.
Study design: Cross-sectional study to evaluate the association between the type of graduate school, the score class obtained at the admission test, the location of the graduate school and the location of the degree.
Methods: The evaluation of the association between site of specialty admission and degree, score at degree and score at admission test was performed by multiple correspondences analysis (MCA). Then, through a logistic regression model, the Odds Ratios (OR) and the respective confidence interval with 95% (95%CI) confidence level of the covariates on the probability of remaining in the same degree site, or in the same region or in the same geographical area, were estimated.
Results The highest admission score and the highest age are significantly and independently associated with the probability of choosing, as a graduate school location, the same location where the degree was obtained.
Conclusion: In conclusion, the choice of the course and the location of the specialty course is made in most cases, taking into account the score made in the exam, based on the location where you attended the course of study in medicine and surgery.

Keywords: Medical Education; Competition; Young medical doctor; Medical specialities; Career choice.

BACKGROUND

In Italy, the issue specialty postgraduation degree in medical or surgical field is subject to the possession of a Degree in Medicine and Surgery and the qualification to practice the medical profession, as well as passing the entrance exam for access to specialty schools, and the completion of the related specialized training course.

Initially, access to the Schools of Specialty for medical doctors was governed by Legislative Decree no. 369 of 17 August 1999 [1], according to which the admission tests were held locally, on the same date each of the three areas (Medical, Surgical, Services) to which each specialty school belongs, with contents defined at national level, according to a prepared calendar. The selection boards were set up at local level, according to predetermined criteria and the applications administered were chosen from a public national database accessible to all. In addition to the score obtained in the entrance tests, each candidate was awarded a score relating to the degree mark and the curriculum of studies.

Since 2017, the procedures for admission of

DOI: 10.54103/2282-0930/19997
Accepted: 20th July 2023

doctors to specialty schools have been governed by Ministerial Decree no. 130 of 10 August 2017 [2], according to which the Medical specialty Schools are accessed through an annual test for qualifications and exams, in which the available places are indicated when the exam date is announced, divided for each school, the topics of study on which the questions are prepared no longer extrapolated from a public national database, the criteria for assigning the score, the calendar, the duration and the methods of carrying out and correcting the exam, as well as the application instructions, of a technical-IT nature, on the methods of administering the questions and correcting them, necessary to guarantee their reliability, transparency and uniformity.

The entrance test is mainly composed of questions related to the evaluation, within predefined mono and / or interdisciplinary scenarios, of clinical, diagnostic, analytical, therapeutic, and epidemiological data.

The questionnaire was entrusted to the Ministry, with the technical-operational support of Cineca, an Italian non-profit inter-university consortium, which, for this purpose, can make use of subjects with proven competence in the field.

Once the exam has been completed, the Ministry of Education, University and Research draw up a single national ranking, containing the overall score achieved by each candidate. In case of an equal score, the candidate who obtained the highest score in the exam prevails, in case of further parity, the youngest candidate.

Choosing a postgraduate career path is an important choice that is often difficult to change once in specialist training [3,4,5]. Career choices made by students shape the human resources landscape in healthcare, and a better understanding of the career choice process can help create a better match of students' preferences with specialist needs [6,7,8].

In the present study we wanted to analyze the possible factors that determined the choice of a given graduate school and the location where to attend it in the year 2021.

MATERIALS AND METHODS

The study was a cross-sectional on the results of the available list of participants to the selection. Data have been made available by Universitaly, that is the official web portal of MUR (Ministry of University and Research) that provide official rankings and allocations. Graduation site was obtained by subscription data of the National Federation of Physicians and Surgeons (Italian acronym FNOMCEO).

The movement of the student specializing between the place of graduation and the location of the graduate school was analyzed in relation to the score obtained on the admission test, age, gender and the presence or absence of the school chosen in the place of graduation. This shift has also been assessed at regional level. Then, through a logistic regression model, the Odds Ratios (OR) and the respective confidence interval with 95% (95%CI) confidence level of the covariates on the probability of remaining in the same degree site, or in the same region or in the same geographical area, were estimated.

The generalized linear models were used to compare, both by degree location and by chosen graduate school, the average scores achieved at the admission test between those who moved and those who did not move. The presence of the school in the place of graduation was considered as a possible modifier of effect. Multiple comparisons have been adjusted according to Tukey.

The evaluation of the multiple association between the type of graduate school, the score class obtained at the admission test, the location of the graduate school and the location of the degree was carried out through the application of the multiple correspondences analysis (MCA, multiple correspondence analysis, and the consequent creation of the Burt table. Scoring classes were identified based on percentiles of the score distribution (<5%, 5%-25%, 25%-50%, 50%-75%, 75-95%, >95%). The inertia of the first two dimensions identified by the MCA is the index for the evaluation of the variability explained by the association between the characteristics. The coordinates of the variables on the first two dimensions were represented graphically to be able to evaluate the associations between the characteristics of the variables.

The size and inertia of each feature were subsequently used as quantitative variables for the identification of aggregated characteristics which was carried out by means of a cluster analysis with the centroid method. In view of the number of graduate schools (in total = 50), as well as assuming that school locations, degree locations and scoring class can aggregate around each school, an approach was carried out with a high number of clusters, over 30, subsequently reduced to 25, in order to avoid clusters with isolated characteristics (clusters composed of a single point). To this end, the distance between the clusters and the loss of variability explained as the clusters decreased were considered, establishing not to fall below 95% of explained variability.

Data were analyzed by SAS software for PC and a $p<0.05$ was set for statistical significance.

RESULTS

The admission test to graduate schools was supported by 19442 graduates in medicine and surgery, however the analysis was carried out on 19269 participants; 173 participants were removed because of missing values on FNOMCEO data related to site of graduation. The average age (ds) is 28.7 years (4.7), of which 57.9% (11148/19269) male

and 42.1% female. A part of the participants appears as "fallen", that is, they have not made any choice or have not been able to obtain the desired choices: they are 3313 graduates, equal to 17.2% of all participants; have an average age (ds) of 29.9 years (5.7).

Among those entering graduate school, 4.2% (666/15968) have a score below 47 (percentile class between the minimum and 5th) points, 17.9% (2869/15968) have a score between 47 and 67 (percentile class 5th-25th), 22.7% (3617/15968) have a score between 67 and 80 (percentile class 25th-50th), 26.3% (4202/15968) have a score between 80 and 95 (percentile class 50°-75°), 23.1% (3685/15968) have a score between 95 and 114 (percentile class 75°-95°), 5.8% (921/15968) have a score between 114 and 138 (percentile class 95°-Maximum). For a residual 0.05% (8/15968) errors are found in the score which is therefore not analyzable.

57.9% (9238/15968) of those who enroll in a graduate school, change university location with respect to the degree location.

The highest admission score and the highest age are significantly and independently associated with the probability of choosing, as a specialty school location, the same location where the medical degree was obtained.

The choice not to change the location of graduate school with respect to the degree location is significantly associated with having achieved a higher score on the admission test and the higher age, respectively OR 1,007 for each additional point (95% CI 1,003-1,006) and OR 1,050 for each additional year (95% CI 1,041-1,059). When the location shift is evaluated between regions, in addition to the score and age, also the gender and the presence of the school within the same university location of graduation are significantly associated with a greater probability of choosing a school within the same region where the degree site is located: test score, OR 1,007 for each additional point (95% CI 1,005-1,008), Age, OR 1,051 for each additional year (95%CL 1,042-1,060); female (F vs M: OR 1.072, 95% CI 1.003-1.146), presence school in the degree (yes vs no: OR 3.801, 95% CI 3.400-4.250).

Tables 1 and 2 show the average scores per degree location and by type of graduate school, estimated in the group of those who move to a place of specialty other than the degree one and in the group of those who do not move. The relevant cases concern on the one hand graduates in Bologna and Palermo: those who remain in the same location have an average score significantly higher than those who choose other locations; on the other hand, the graduates in Foggia and Sassari for whom the highest score belongs to those who choose a school location different from the degree one. In relation to the School of Specialty, those who achieve a higher score on the admission test tend to choose the School in the same place of graduation, but the difference in score between those who change location of the university in which they graduated compared to those who do not change it, is statistically significant only for Schools in Radiodiagnostics, Internal Medicine, Hygiene and Pathological Anatomy.

The multiple correspondence analysis (figures 1, 2, 3 and 4) allows to highlight any associations between the score, locations and type of specialty school chosen. The graphs show the arrangement of the characteristics according to the first two dimensions that explain 23.2% of the variability. A decreasing trend can be observed in the score from left to right and the schools that orbit around the points corresponding to the scoring class on the test represent the type of school preferred over the score; another trend from left to top (II quadrant) to bottom right (IV quadrant) shows the arrangement of the degree seats and specialty locations, according to a north-south axis.

The analysis of the clusters identified the 25 groups that are shown in 4 different figures. Cluster formation allows for a better interpretation of the associations between features. Cluster 6, the first to be encountered from the left (Figure 1), contains high scores (above the 95th percentile), the graduate school in dermatology, the headquarters of the San Raffaele University School in Milan. This aggregation means that those who obtained the highest scores were able to choose first, and preferentially choose as a school the one in dermatology (regardless of the location) and as the seat of the school the San Raffaele University of Milan (regardless of the school). In the immediate vicinity (Figure 1) are clusters 2, 7 and 8. Cluster 2 mainly aggregates graduate school locations that are located in central and northern Italy. Cluster 7 mainly aggregates the scoring class between the 75th and 95th percentile and some types of specialty school obviously very coveted (Plastic Surgery, Neurology, Ophthalmology, Cardiology, Endocrinology, Diseases of the Digestive System and Pediatrics) and. Cluster 8, on the other hand, mainly contains graduate school locations at the universities of Milan.

The association of the scoring class with the locations is less clear: these clusters mainly contain degree and school locations in the central north, without a particular association between the type of school and the school location or flows between locations being highlighted, rather it could be the confirmation that the location chosen for the specialty is the same or close to the degree location.

Clusters 1 and 4 (Figure 1) are located at the bottom and rather detached from the other points (also compare the provisions in figures 2, 3 and 4); these represent the university campuses of Campania, highlighting the tendency to remain in the graduation region or to move within the region. This feeling is also confirmed in the points that belong to clusters 3, 5 and 10. These clusters are located around the point that represents the scoring class between the 5th and 25th percentiles.

Clusters 11 to 14 (Figure 2) are very close to each other (remember that the proximity of the points is not necessarily an indication of the strength of the association). Cluster 13 is characterized mainly by types of graduate schools and contains the scoring class between the 25th percentile and the median. Clusters 11, 12 and 14 mainly group together degree and school campuses: in particular, clusters 12 and 14 group the Roman universities.

Cluster 15 (Figure 3) contains the class of points between the median and 75th percentile, Northeast school and graduate locations, and graduate schools such as vascular surgery, pediatric surgery, psychiatry.

The last clusters, from 17 to 25 (Figure 4), represent almost regional aggregations of locations, both undergraduate and graduate school.

DISCUSSION AND CONCLUSIONS

The analysis of multiple matches, before being carried out to date for this topic allows to evaluate the association between the type of school chosen, the result of the test and the location of the preferred school.

The analyses presented did not highlight mobility flows. The idea that higher rankings choose more prestigious locations is not fully supported by the data. The result that has been obtained suggests that there is little mobility between the locations: the competitors prefer to stay in their own graduation site or in nearby locations. In addition, what emerges from the data is that doctors who achieve high scores and who have graduated from university campuses in the south prefer to move to universities in the north. However, the mobility flows highlighted may depend on the unavailability of places for graduates of the location rather than on the ranking and prestige of the school location [9]. Another hypothesis could rise from incurring costs related to mobility, our data couldn't help to analyze deeply this aspect: probably the mobility has taken place from the matriculation at first year of the graduation course, then once settled the student prefer to not move; on the other hand, those who cannot afford the mobility have chosen a site near home, even if they didn't realize their main ambition. Further studies are necessary to explore these hypothesis, with questionnaires to students.

However, the study has limitations. First, the analysis was carried out only on the choice made, not being aware of what the initial preferences of the competitors were. This information is essential to better interpret the association between score and type of school. Although it is quite evident that the highest score corresponds to schools such as dermatology, plastic surgery and ophthalmology, there is no information to say that the highest rankings (highest test points) choose the schools. It cannot also be excluded that among the highest rankings there are candidates who have indicated other schools than those previously mentioned.

In various research the chosen specialty was similar to those observed in this research [10]. More in-depth analyses, however, have shown how it is possible to associate the choice with the expectations of personal and professional fulfilment, with a view to a good life-work balance [11,12].

Specializations related to public health and general practice, as observed in our analysis, are less preferred [10], and often associated with the female sex [13,14].

These observations open to the further observation to the planning of specialty with respect to the real needs of the Italian healthcare system, anyway the choice of young doctors does not always take this last aspect into account [9]. However, the data available to us do not allow us to further investigate this aspect arising from the literature.

In an exploratory analysis, but not shown, carried out considering the evaluations of the Censis surveys (data not shown), it associates the top ten positions with the northern universities to which mainly northern graduate students did access, but it was not possible to associate types of schools.

It should also be noted that as of 27 January 2022 the number of unassigned contracts, compared to 18847 contracts announced 2539 contracts were not covered, equal to 13.5%. As can be seen from Table 3, however, there is great inhomogeneity: there are specializations that are in fact completely occupied, compared to very high percentages of unoccupied places in other specializations.

To the 2539 contracts 276 training contracts already abandoned must be added and therefore can no longer be awarded. The provisional result is therefore 15% of unoccupied posts. Solutions to avoid this waste of resources could be in a guided choice during the degree course [9,14,15], to increase awareness of their own propensity to certain specialty. Some authors have studied the personality trait of student (extraversion, conscientiousness, agreeableness, neuroticism, openness) [16], but these aspects couldn't be evaluated through the analysis of the simple provisional ranking.

In conclusion, the choice of the course and the location of the specialty course is made in most cases, taking into account the score made in the exam, based on the location where you attended the course of study in medicine and surgery. However, taking into account that the distribution of places in the different specialty should have the purpose of training a given number of specialists having in mind the demand and needs of the Italian health system, it emerges as a criticality that the specialty in which the demand does not meet the supply are the same in which there is greater shortage: for the specialty in emergency medicine compared to 1189 banned posts, only 665 were occupied, equal to 44.1%.

With this work, we wanted to explore factors involved in the choice of the specialty course and give an analysis that could help to give direction during the training of the Italian medical doctor accounting for the need of the sustainability of Italian national health system.

REFERENCES

1. Legislative Decree no. 369, 17 August 1999, Implementation of directive 93/16/EEC on the free movement of doctors and mutual recognition of their diplomas, certificates and other qualifications and of directives 97/50/EC, 98/21/EC, 98/63/EC and 99/46/EC amending directive 93/16/EEC)). pubblicato nella Official Bullettin of Italian Republic n. 250, 23 ottobre 1999 – Ordinary Supplement n. 187, (Amendment Official Bulletin. n. 44, 23 febbraio 2000)
2. Ministry of Education University and Research, Decree 10 August 2017, n. 130 Regulation of admission test to schools of specialization in medicine and surgery, Official Bullettin of Italian Republic n. 208, 06/09/2017.
3. Ladha F.A., Pettinato A.M. and Perrin A.E. Medical student residency preferences and motivational factors: a longitudinal, single-institution perspective. BMC Med Educ 2022;22(1):187.
4. Douglas Jones M.J., Yamashita T., Ross R.G. and Gong J. Positive predictive value of medical student specialty choices, BMC Med Educ 2018;18:33.
5. Woolf K., Elton C. and Newport M. The specialty choices of graduates from Brighton and Sussex Medical School: a longitudinal cohort study. BMC Med Educ 2015;15:46.
6. Di Silverio P., Conte A., D'Arenzio M., Ragazzo F., Atzeni D., Fodale A, [Specialistic Medical Educationa and triennal plannig] Formazione Specialistica e Programmazione triennale, Dirigenza Medica, n.1, 2018.
7. Di Silverio P., Conte A., D'Arenzio M., Ragazzo F., Atzeni D., Fodale A., [An algorythm to determine number of physicians required] Medici, un algoritmo sui fabbisogni, Il Sole 24 Ore Sanità, n.3, 2018.
8. Detsky A.S., Gauthier S.R., Fuchs V.R. Specialization in Medicine. How Much Is Appropriate? JAMA. 2012; 307; 5.
9. Errichiello C., De Gregorio MG. The new exam for admission to Specialty Schools in medicine and surgery: lights and shades. Il nuovo concorso nazionale italiano di accesso alle scuole di specializzazione: luci e ombre. Giornale di Tecniche Nefrologiche e Dialitiche 2015; (3):195-197.
10. Norredam M., Album D. Prestige and its significance for medical specialties and diseases. Scand J Public Health, 2007; 35: 655–661.
11. Sarikhani Y., Ghahramani S., Bayati M., Lotfi F. and Bastani P. A thematic network for factors affecting the choice of specialty education by medical students: a scoping study in lowand middle-income countries, BMC Med Educ; 21(1):99.
12. Hohf-Nova M., Hun-Pacheco R., Muñoz-Bustos D., Soto-Carriel A. and Pérez-Villalobos C. When it is time to decide: factors associated to the choice of a medical specialty. Rev Med Chile 2021,149:1352-1359.
13. Asaad M., Zayegh O., Badawi J. et al. Gender differences in specialty preference among medical Students at Aleppo University: a cross-sectional study, BMC Med Educ 2020;20:184.
14. Morte K., Nelson D., Marenco C., Lammers D., De Barros M., Bader J., and Bingham J., Gender Differences in Medical Specialty Decision Making: The Importance of Mentorship. J Surg Res 2021; (267): 678–686.
15. Hao K.A., Fu S., Islam S., Larson S.D., Mustafa M.M., Petroze R.T., and Taylor J.A. Medical Student Career Choice: Who Is the Influencer?. J Surg Res 2022; (272): 9–16.
16. Kwon O.Y. and Park S.Y. Specialty choice preference of medical students according to personality traits by Five-Factor Model, Korean J Med Educ 2016; 28(1): 95-102.

TABLES

Table 1. Average scores achieved by aspiring postgraduates in relation to the degree location and whether they have changed university location for graduate school.

Graduation Location	specialty **in another location**	specialty **same location**	**adj p-value**
SS	81,51±1,74	67,16±2,05	**0,0003**
FG	75,53±1,9	64,09±1,83	**0,0302**
AQ	80,03±1,45	69,88±1,96	**0,0555**
FE	82,1±1,29	72,5±2,14	0,1663
PV	86,47±1,2	78,39±2,11	0,5437
VA	85,84±1,75	77,98±2,6	0,9941
UNIMOL	78,53±1,89	70,89±7,19	1,0000
SI	84,59±1,34	77,25±2,21	0,9238
UNPIE	89,56±2,03	82,51±3,18	1,0000
VR	92,23±2,13	86,34±1,91	1,0000
PG	81,92±1,2	76,84±1,82	0,9989
RM_TV	81,41±1,24	76,89±1,79	1,0000
CA	81,94±1,77	77,82±1,35	1,0000
UD	90,14±2,54	86,07±3,48	1,0000
PD	88,61±1,29	84,88±1,06	0,9994
POLMAR	89,42±1,99	85,81±1,54	1,0000
UNCAM	77,34±0,87	74,27±1,13	0,9998
PI	84,65±1,45	81,93±1,38	1,0000
CZ	75,26±1,39	72,63±1,69	1,0000
FI	87,95±1,3	85,58±1,31	1,0000
NA_FED	84,55±1,09	82,86±1,16	1,0000
CH	79,26±1,23	78,17±1,87	1,0000
GE	83,53±1,62	82,6±1,22	1,0000
RM_SAP	81,42±0,69	81,49±0,83	1,0000
RM_BIO	84,41±1,57	84,74±3,27	1,0000
BA	74,14±1,29	74,99±1,13	1,0000
PR	81,73±1,28	82,67±1,89	1,0000
TO	85,02±1,42	86,77±0,93	1,0000
MI_BIC	89,2±1,79	91,25±2,49	1,0000
ME	68,17±1,12	70,29±1,43	1,0000
BS	82,26±1,66	84,42±1,46	1,0000
MO_RE	77,95±2,05	80,54±1,83	1,0000
CT	75,72±0,97	79±1,22	1,0000
TS	83,74±1,74	88,63±2,27	1,0000
MI_SRAF	92,81±1,5	98,93±2,68	1,0000
MI	87,82±1,23	94,07±1,17	0,2066
SA	80,39±1,21	87,25±2,24	0,9748
BO	82,85±1,1	90,31±1,24	**0,0092**
PA	71,67±0,82	79,57±1,04	**<,0001**
RM_CAT	85±1,19	93,05±1,66	0,0999
MI_HUM	91,2±2,47	104,1±5,28	0,9998

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Table 2. Average scores achieved by aspiring postgraduates in relation to the type of specialization school chosen and whether they have changed university location for the specialization school.

Postgraduate School	specialty **in a location other than Bachelor's degree**	specialty **in the same degree location**	**adj p-value**
MedTer	46,79±13,99		
Stat	71,36±3,88	47,11±6,26	0,7034
Farm	55,24±3,39	50,12±2,34	1,0000
Alim	71,88±2,4	68,11±2,41	1,0000
CardCh	80,13±1,79	76,75±2,86	1,0000
ChPla	109,4±1,39	106,53±2,81	1,0000
Mdig	101,55±1,2	98,81±1,35	1,0000
Oto	94,62±1,26	93,01±1,61	1,0000
Micr	55,11±3,5	53,82±2,56	1,0000
Minf	75,69±1,05	74,51±1,15	1,0000
Gen	66,87±2,51	66,89±2,65	1,0000
MedSp	89,83±1,98	90,04±2,56	1,0000
Reum	93,51±1,63	93,82±1,95	1,0000
MedUrg	63,72±0,92	64,11±0,88	1,0000
ChVas	76,69±1,55	77,24±2,07	1,0000
Audio	60,8±3,06	61,46±4,04	1,0000
ChMax	85,12±2,16	85,94±3,3	1,0000
Npi	82,84±1,04	83,93±1,43	1,0000
Onc	83,24±0,99	84,33±1,19	1,0000
Derm	111±1,41	112,24±1,74	1,0000
Ema	86,14±1,17	87,41±1,32	1,0000
ChGen	70,88±0,76	72,72±0,81	1,0000
ChPed	80,58±1,87	82,53±3,88	1,0000
Allerg	85,99±1,81	88,12±2,01	1,0000
Ort	83,55±0,78	86,07±1,03	1,0000
Anest	68,95±0,48	71,59±0,55	0,2163
MedNuc	55,9±2,6	58,62±2,75	1,0000
Ped	101,29±0,62	104,15±0,76	0,8755
MedLeg	83,39±1,32	86,33±1,62	1,0000
Mresp	80,87±0,99	83,82±1,1	1,0000
Mcard	106,84±0,66	110,13±0,8	0,6737
PatCl	53,88±2,11	57,32±1,66	1,0000
Neur	102,39±0,98	105,83±1,21	1,0000
Oft	102±1,06	105,65±1,47	1,0000
Psi	79,46±0,69	83,34±0,86	0,4137
Gin	94,39±0,73	98,36±1,03	0,7618
Rad	53,04±2,04	57,06±2,05	1,0000
Radiagn	83,28±0,64	87,33±0,75	**0,0462**
MedInt	77,81±0,66	82,37±0,66	**0,0008**
MedLav	71,79±1,3	76,55±1,38	0,9981
Uro	81,87±1,07	86,9±1,36	0,9356

MedFis	71,81±1	76,96±1,23	0,7353
NeurCh	90,52±1,55	95,71±2,37	1,0000
Ger	67,56±0,99	72,89±0,88	0,0924
Endo	98,85±1,18	104,38±1,38	0,8768
MedCom	53,79±2,44	59,63±2,7	1,0000
Ig	60,53±0,77	66,38±0,78	**0,0001**
Nefr	69,15±1,03	75,96±1,43	0,2030
ChTor	60,98±2	71,66±2,65	0,7699
AnaPat	62,93±1,98	74,68±1,63	**0,0153**

Table 3. Contracts not awarded as of 26 January 2022

	Banned contracts	**Unassigned contracts**	**Abandoned contracts**	**% unallocated + abandoned**
Thermal medicine	4	3	1	100,0%
Microbiology and virology	154	104	2	68,8%
Clinical Pathology and Clinical Biochemistry	347	228	0	65,7%
Emergency medicine	1189	665	11	56,9%
Pharmacology and Clinical Toxicology	119	63	0	52,9%
Health Statistics and Biometrics	39	20	0	51,3%
Nuclear medicine	112	55	1	50,0%
Radiotherapy	186	90	3	50,0%
Pathological anatomy	216	80	5	39,4%
Medical genetics	96	31	1	33,3%
Community and Primary Care Medicine	89	27	1	31,5%
Audiology and phoniatrics	46	12	2	30,4%
Cardiac surgery	108	16	12	25,9%
Thoracic Surgery	99	22	3	25,3%
Anaesthesia Resuscitation	2155	358	42	18,6%
Food science	87	14	2	18,4%
Infectious and Tropical Diseases	393	59	11	17,8%
General Surgery	817	103	29	16,2%
Paediatric surgery	80	11	0	13,8%
Nephrology	341	40	5	13,2%
Vascular Surgery	147	14	5	12,9%
Hygiene and preventive medicine	809	78	13	11,2%
Geriatrics	545	55	4	10,8%
Internal medicine	1151	89	20	9,5%
Diseases of the respiratory system	418	33	5	9,1%
Haematology	287	19	7	9,1%
Urology	309	8	15	7,4%
Medical Oncology	378	23	5	7,4%

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Physical and rehabilitation medicine	369	22	5	7,3%
Maxillofacial surgery	69	5	0	7,2%
Sports and exercise medicine	98	7	0	7,1%
Occupational medicine	244	14	2	6,6%
Rheumatology	142	9	0	6,3%
Child neuropsychiatry	304	16	2	5,9%
Allergology and clinical immunology	117	4	2	5,1%
Orthopaedics and traumatology	580	16	13	5,0%
Gynaecology and Obstetrics	623	21	10	5,0%
Neurosurgery	125	2	4	4,8%
Forensics	207	7	2	4,3%
Radio diagnostics	985	27	10	3,8%
Diseases of the digestive system	267	10	0	3,7%
ENT	218	6	2	3,7%
Psychiatry	759	21	6	3,6%
Neurology	362	10	2	3,3%
Endocrinology and metabolic diseases	251	2	4	2,4%
Reconstructive and aesthetic plastic surgery	133	2	0	1,5%
Paediatrics	973	11	3	1,4%
Ophthalmology	281	2	1	1,1%
Diseases of the cardiovascular system	849	5	3	0,9%
Dermatology and venereology	170	0	0	0,0%
Total	**18847**	**2539**	**276**	**14,9%**

Table 4. Award of the first 1000 contracts

Specialty	**N.**
Diseases of the cardiovascular system	277
Paediatrics	139
Dermatology and venereology	79
Neurology	69
Ophthalmology	58
Reconstructive and aesthetic plastic surgery	39
Gynaecology and Obstetrics	38
Endocrinology and metabolic diseases	36
Diseases of the digestive system	32
Internal medicine	26
Psychiatry	22
Radio diagnostics	21
Haematology	19
Orthopaedics and traumatology	19

Anaesthesia Intensive Care And Pain Intensive Care	18
ENT	12
Infectious and Tropical Diseases	10
Neurosurgery	9
General Surgery	8
Medical Oncology	8
Rheumatology	8
Urology	7
Allergology and clinical immunology	6
Pathological anatomy	6
Geriatrics	5
Cardiac surgery	3
Diseases of the respiratory system	3
Emergency medicine	3
Forensics	3
Child neuropsychiatry	3
Paediatric surgery	2
Vascular Surgery	2
Sports and exercise medicine	2
Physical and rehabilitation medicine	2
Maxillofacial surgery	1
Thoracic Surgery	1
Medical genetics	1
Hygiene and preventive medicine	1
Nephrology	1
Health Statistics and Biometrics	1

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Figure 1. Scatter plot of data point for clusters 1 -10. Data Point are university of graduation (GR_), university sites of specialization school (SP_), the specific school and the result at the selection exam. The x-axes is related to a geographical trend in school university from north to south, and to a trend in decreasing result of the exam. The y-axes is related to the graduation university.

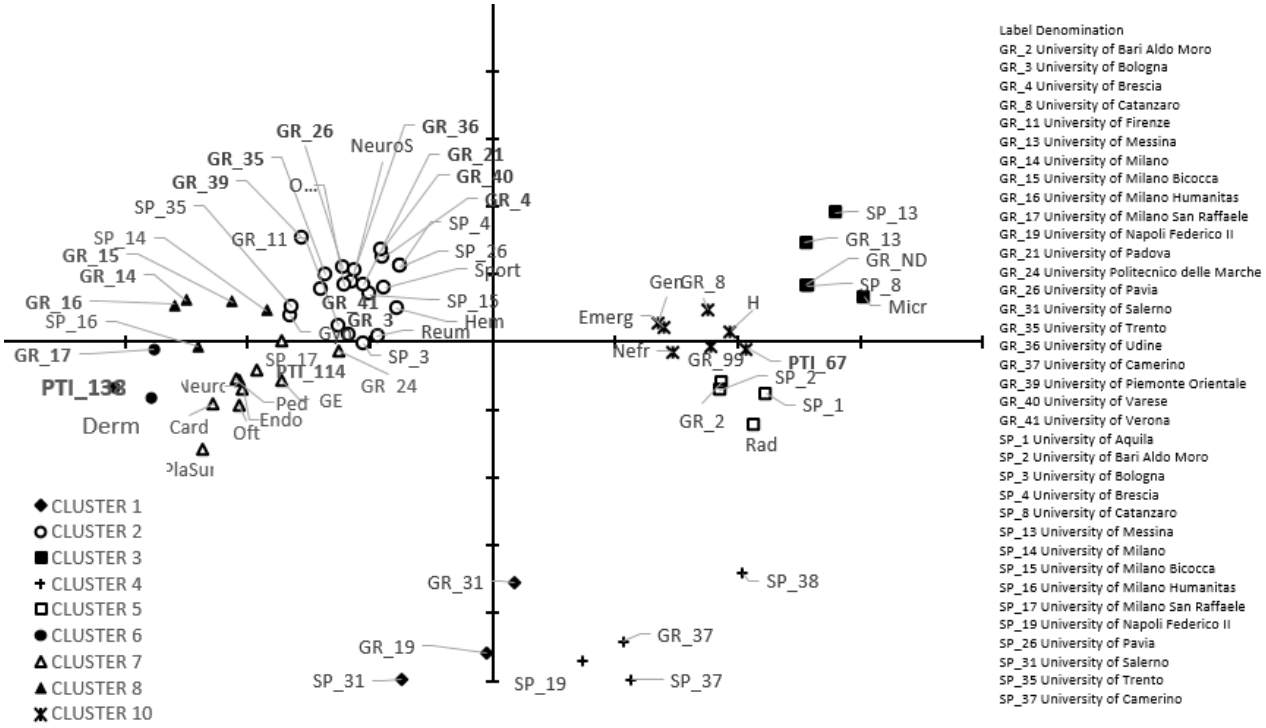
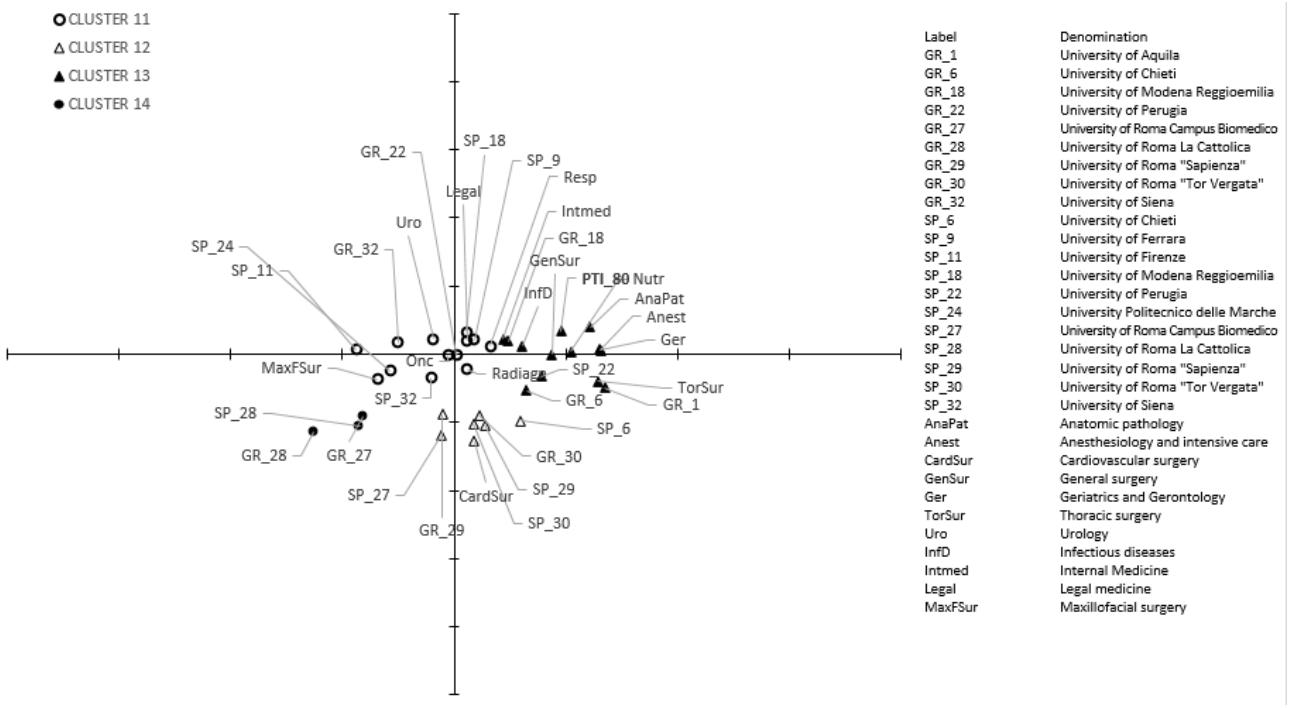


Figure 2. Scatter plot of data point for clusters 11 -14. Data Point are university of graduation (GR_), university sites of specialization school (SP_), the specific school and the result at the selection exam. The x-axes is related to a geographical trend in school university from north to south, and to a trend in decreasing result of the exam. The y-axes is related to the graduation university.



Label	Denomination
GR_1	University of Aquila
GR_6	University of Chieti
GR_18	University of Modena Reggioemilia
GR_22	University of Perugia
GR_27	University of Roma Campus Biomedico
GR_28	University of Roma La Cattolica
GR_29	University of Roma "Sapienza"
GR_30	University of Roma "Tor Vergata"
GR_32	University of Siena
SP_6	University of Chieti
SP_9	University of Ferrara
SP_11	University of Firenze
SP_18	University of Modena Reggioemilia
SP_22	University of Perugia
SP_24	University Politecnico delle Marche
SP_27	University of Roma Campus Biomedico
SP_28	University of Roma La Cattolica
SP_29	University of Roma "Sapienza"
SP_30	University of Roma "Tor Vergata"
SP_32	University of Siena
AnaPat	Anatomic pathology
Anest	Anesthesiology and intensive care
CardSur	Cardiovascular surgery
GenSur	General surgery
Ger	Geriatrics and Gerontology
TorSur	Thoracic surgery
Uro	Urology
InfD	Infectious diseases
Intmed	Internal Medicine
Legal	Legal medicine
MaxFSur	Maxillofacial surgery

Figure 3. Scatter plot of data point for cluster 15. Data Point are university of graduation (GR_), university sites of specialization school (SP_), the specific school and the result at the selection exam. The x-axes is related to a geographical trend in school university from north to south, and to a trend in decreasing result of the exam. The y-axes is related to the graduation university.

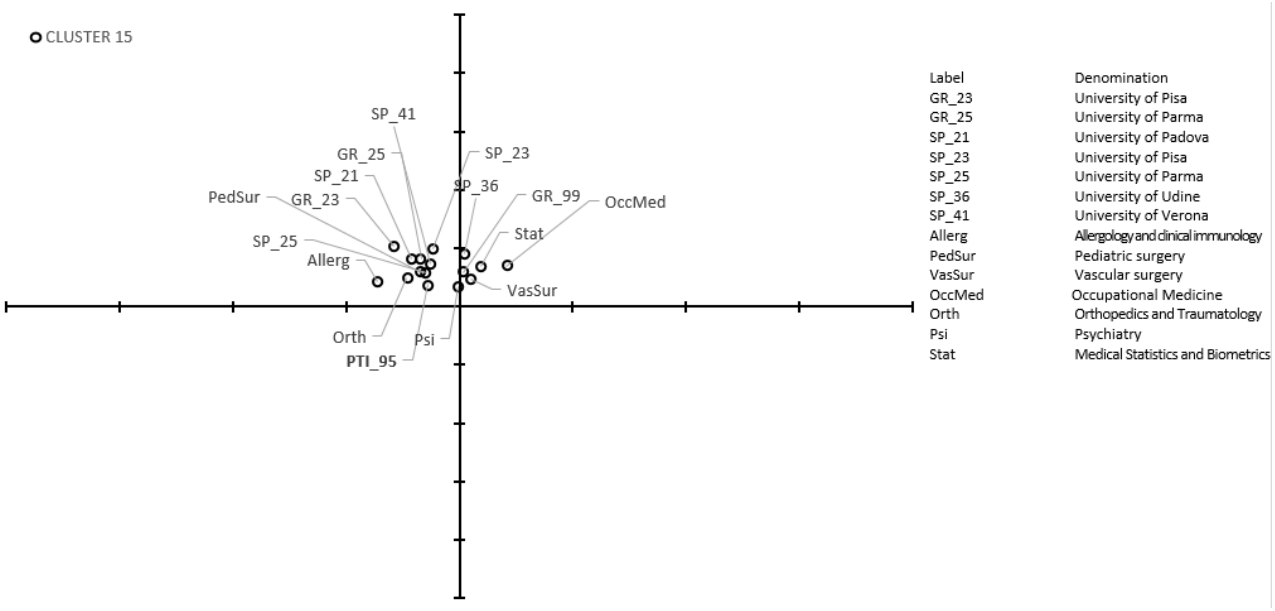
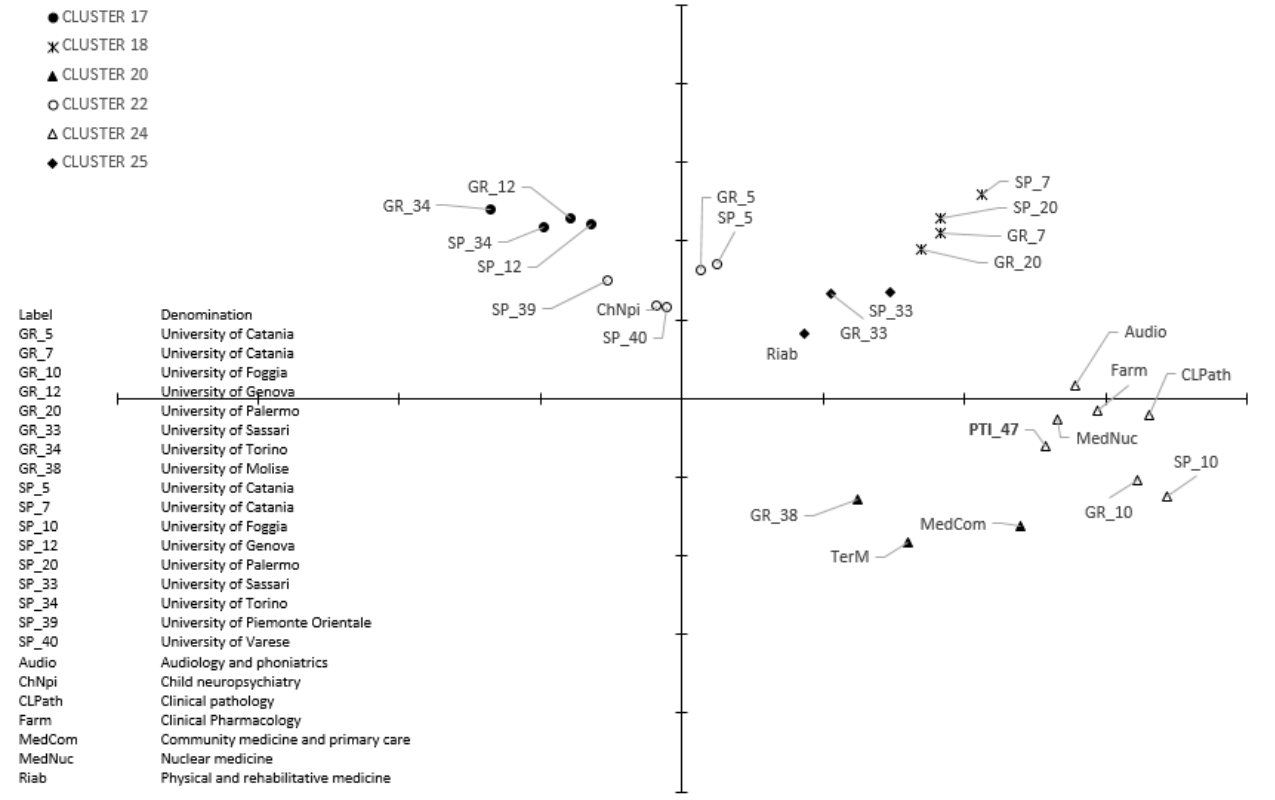


Figure 4. Scatter plot of data point for clusters 17 - 25. Data Point are university of graduation (GR_), university sites of specialization school (SP_), the specific school and the result at the selection exam. The x-axes is related to a geographical trend in school university from north to south, and to a trend in decreasing result of the exam. The y-axes is related to the graduation university.



ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Stem-Skilled Parents and Autism Spectrum Disorder in Offspring: A Case-Control Study

Marco Valenti[1,2], Margherita Attanasio[1], Ilenia Le Donne[1], Antony Bologna[1], Francesco Masedu[1], Monica Mazza[1,2]

(1) Department of Applied Clinical Sciences and Biotechnology, University of L'Aquila, Italy
(2) Reference Centre for Autism of the Abruzzo Region, L'Aquila, Italy

CORRESPONDING AUTHOR: Professor Marco Valenti, MD, University of L'Aquila, Department of Applied Clinical Sciences and Biotechnology, Via Vetoio, Edificio De Meis, L'Aquila 67100, Italy; Director of the Regional Reference Centre for Autism, L'Aquila, Italy. Email: marco.valenti@univaq.it, ORCID ID 0000-0001-9043-3456

SUMMARY

Autism spectrum disorder (ASD) is a neurodevelopment disorder characterised by a range of deficits in two specific domains: social communication and social interaction and repetitive patterns of behaviour. Several studies have explored the link between ASD and STEM (science, technology, engineering and mathematics, or other mathematics-grounded disciplines), but results are still uncertain. Objective of the study was to estimate the potential role of systemising abilities in parents as a risk factor for ASD in the offspring, using the achievement of a degree in STEM disciplines as a proxy characteristic of the exposure. There were 1,316 participants overall. There were 658 incident consecutive cases of definite ASD, diagnosed in a Reference Centre for ASD in Italy, from 2001 to 2020. The main exposure variable was parental education level. The risk of ASD in the offspring associated with the main exposure variable and the exposure covariates (e.g. use of neurotropic drugs during the first trimester of the mother's pregnancy, perinatal outcomes of participants and/or preterm birth) was studied by using conditional logistic regression analysis. In addition, we carried out a mediation analysis to investigate whether and the extent to which covariates significantly associated with ASD risk mediate the relationship between parental education level and ASD in offspring. A STEM degree in parents was significantly associated with risk of ASD in offspring (OR 1.43, 95% CI 1.03-2.54). Familiarity was weakly associated with the risk of ASD (OR 1.33, 95% CI 1.00-1.66) and is the stronger mediator (PME 28%). Sensitivity analysis did not show deviations related to gender or ASD level.
Our study moves in the direction of confirming the risk of occurrence of ASD in the offspring of parents with elevated systemising abilities.

Keywords: Autism spectrum disorder (ASD); STEM disciplines; systemising abilities; ASD risk in offspring.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The literature supports association between ASD and STEM disciplines (science, technology, engineering and mathematics, or other mathematics-grounded disciplines): this study gives a sound contribution to the debate and, overall, contributes to enrich knowledge and eventually leads to advances in information on the aetiology of ASD.
- The design and the sample size of the study allow well-powered conclusions, although larger samples are required to investigate more possible confounders and effect modifiers
- The main limitation is the lack of variables such as environmental or dietary exposures or genetic profiling. In addition, a prospective study following STEM graduates over time and monitoring offspring would be optimal, although inefficient.

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopment disorder that emerges in early childhood, characterised by a range of deficits in two specific domains: social communication and social

DOI: 10.54103/2282-0930/20742
Accepted: 27th July 2023

interaction and repetitive patterns of behaviour [1]. ASD is considered a medical condition that gives rise to disability as well as an example of human neurological variation ('neurodiversity') that defines a person's identity, with cognitive assets and challenges [2-3].

The epidemiological interest in ASD has increased because of its growing awareness in Western countries, documented in terms of both clinical research and public health attention [4]. Epidemiological research has highlighted the impact of ASD all over the world [5]. Regarding the United States, the surveillance system network of 11 states [6] estimated an overall prevalence of ASD of 1.68% (2.66% males and 0.66% females, with a male to female ratio of 4:1). The average prevalence of ASD in the European Union program is reported as 1.0% [7]. Indeed, in the last decades ASD diagnoses have at least doubled: in addition to a true increase in prevalence, the literature proposes alternative explanations, including changing diagnostic criteria, different methods of ascertainment, inhomogeneous protocols of diagnosis, research protocols, environmental components, cultural factors or awareness in recent years [8]. In any case, research is attempting to understand the aetiology of ASD, in terms of the epigenetic, neurobiological, genetic, neurological and hormonal factors at the base of this complex condition. There is no agreement about the causes of autism, and the aetiopathogenetic factors of ASD remain unknown. According to Stubbs and collaborators [9] and to Lai et al. [2] , the interactions between genetic and environmental factors are thought to contribute to its causes. There have been numerous studies showing that there is increased risk of neuropsychiatric disorders in offspring due to prenatal and perinatal environmental exposures [10]: nonetheless, genetic factors likely have an important role in ASD development. Specific thinking style [11] and impaired ability in reading intentions of others (theory of mind), which can be considered pathognomonic of ASD, as well as uncommon calculation abilities in a variable proportion of people with ASD, are characteristics strongly correlated with the brain attitude in systemising. This suggests that some of the genes for autism may be regulators of the systemising abilities of the brain, which can be expressed as a continuum regarding the ability to systemise. Moreover, genes can be expressed in first degree relatives that lead to a talent in systemising, thus individuating a trans-generational 'broader phenotype'.

A good proxy descriptor for systemising abilities is the attitude of parents to the so-called STEM disciplines (science, technology, engineering and mathematics, or other mathematics-grounded disciplines). This view has moved ASD research into an understudied area of interest: is ASD more common in the offspring of parents who show high-level skills in systemising abilities and who reach high education levels in the STEM disciplines? In most analytical studies [12] researchers have focused primarily on rates of autism among offspring of adults in the STEM domain; they suggest an elevated prevalence of autisti99c probands and relatives in STEM-related careers [13-15]. In a similar vein, a population-based study of the tech-heavy San Francisco Bay Area suggests that maternal STEM careers are associated with a higher prevalence of offspring with autism, though the researchers found no paternal STEM career choice or joint effects [16].

Some descriptive studies have explored the issue of ASD and STEM, focusing on the hypothesis that autism rates are higher in regions that have a high concentration of jobs in the STEM domain. Roelfsema et al. [17] reported that the prevalence estimates of ASD in an information technology (IT) area in the Netherlands was at least two- to fourfold higher than in non-IT areas, while the prevalence for the control conditions were similar in all regions. The authors underlined that these results are in line with the idea that in regions where parents gravitate towards jobs that involve strong 'systemising', such as the IT sector, there will be a higher rate of autism among their children. However, it is worth noting that conclusions from descriptive epidemiological studies whose objective is to compare prevalence rates of ASD across different areas are possibly jeopardised by relevant biases, like over- or under-diagnosis of borderline cases due to broader clinical criteria, or different awareness about autism in a cross-cultural perspective.

The literature suggests that many professionals in highly systemised occupations who excel in their fields have undiagnosed high-functioning ASD [18-19]. Baron-Cohen et al. [20] demonstrated that a group of undergraduate students with majors in science and mathematics scored significantly higher on all areas of the Autism Quotient (AQ) scale compared with classmates with other majors. Given the 'broader phenotype' symptoms of ASD seen in some parents of children with ASD, some researchers have proposed that these parents have highly technical and structured occupations in fields such as science, engineering and accounting [21-23]. For example, Jarrold and Routh [12] analysed data from Baron-Cohen et al. [21] and reported that occupations in engineering, accounting, science and medicine were more frequent in fathers of children with ASD. Notably, Windham et al. [16] demonstrated in a population-based study in California that the risk of having a child with ASD was almost twofold greater for mothers in highly technical occupations.

Dickerson et al. [24] found that fathers in health care and finance were more likely to have children with ASD; moreover, joint effects of parental technical occupations were associated with communication and social impairment: the results from this study support that a 'broader phenotype' and a sort of 'assortative mating' in adults with autism-like characteristics might contribute to intergenerational transmission of ASD, thus providing, at least, a minimal basis for new genetic models of autism. Overall, the issue of the association between STEM and ASD remains uncertain, and the literature recommends further studies on this topic.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

OBJECTIVE

Objective of the study was to estimate the potential role of optimizing abilities in parents as a risk factor for ASD in the offspring, using the achievement of a degree in STEM disciplines as a proxy characteristic of the exposure.

METHODS

Design and Conduct

We designed and conducted a retrospective case-control study based on population data from registry and outpatient records. The retrospective design was dictated by the relatively low incidence of ASD among the general population. This case-control study was based on incident cases recruited in a specialised centre in Italy and included in an epidemiological registry from 2001 to 2020. The conduct of the study includes the period 2014-2020.

Study Population

There were 1,316 participants overall. There were 658 incident consecutive cases of definite ASD, diagnosed at the Regional Reference Centre for ASD in L'Aquila, Italy, from 2001 to 2020. Diagnosis was made according to the criteria of the *Diagnostic and Statistical Manual of Mental Disorders IV* [25]– or the *DSM-5* [1] after 2013 – and the cut-off values of the Autism Diagnostic Observation Schedule (ADOS) score, first edition [26] or second edition [27] after 2012. In 352 out of 658 cases (53.5%), the individuals were residents in the administrative area of L'Aquila and 0–18 years old, according to the epidemiological registry of the centre [28]. In 306 out of 658 cases (46.5%), the individuals were in the age range of 3–29 years, not resident in the same area and recruited from the autism outpatient clinic of the centre. Given the specific objective of the study and the design, there was no need to define exclusion criteria. There were 475 male cases (mean age at diagnosis 4.2 years, *SD* 2.1 years, range 2–28 years) and 183 female cases (mean age at diagnosis 4.7 years, *SD* 2.7 years, range 2–29 years). With regard to severity, the DSM-5 classifies cases in three levels (level 1, "requiring support"; level 2, "requiring substantial support"; level 3, "requiring very substantial support"): using the *DSM-5* criteria directly (for cases recruited or followed up after 2013) or by indirect attribution (for cases recruited before 2013), 127 out of 658 cases (19.3%), 98 males and 29 females, were classified as level 1; 304 (46.2%), 230 males and 74 females, were classified as level 2; and 227 (34.5%), 147 males and 80 females, were classified as level 3.

Controls were consecutive subjects who had first access to the same centre for clinical observation and diagnosis during the period of 2014–2020, but who did not meet fulfil the criteria for a formal diagnosis of ASD. Thus, 658 sex- and age-matched (± 1 year) controls (males: n = 475, M_{age} = 4.4, SD_{age} = 2.0, range 3–28 years; females: n = 183, M_{age} = 4.4, SD_{age} = 2.4, range 2–29 years) were recruited. No control subjects received an ASD diagnosis during the conduct of the study; 351 controls remained without any other diagnosis, and 307 received diagnoses including intellectual disability, language disorder, attention-deficit/hyperactivity disorder, specific learning disorder, unspecified neurodevelopmental disorder or schizophrenia spectrum disorders.

Patient and Public Involvement

Participants were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Exposure

Data were obtained both by clinical records and personal interviews of parents or legal tutors and collected through a closed-answer questionnaire. Considering that cases had been diagnosed and recruited in the register since 2001, information for cases diagnosed before 2014 was obtained by parental interview at the first follow-up observation during the conduct of the study (i.e. in the time lapse of 2014–2020). Interviews were conducted by a team of operators trained in data collection. The same questionnaire was also given to the parents of control subjects by the same team of operators at the time of the first access to clinical examination. The main exposure variable was parental education level (primary or lower secondary, upper secondary, degree/master in STEM or other discipline).

Mediator variables

We obtained information on several variables considered to be established risk factors related to ASD and evaluated these factors as potential mediators. The questionnaire included the following key covariates: use of neurotropic drugs during the first trimester of the mother's pregnancy (i.e. continuous use by specific medical prescription: yes or no); perinatal outcomes of participants (low birthweight [< 2,500 g] and/or preterm birth [< 37 weeks of gestation]: yes or no), housing and parents' work as a proxy indicator of socioeconomic status (low, middle and elevated); and familiarity, including both ascendants and collaterals, for ASD or other neurodevelopmental or psychiatric disorder (yes or no).

Ethics Approval

All parents or legal tutors of both cases and

controls agreed to participate in the study. The study was approved by the Ethical Committee of the Local Health Agency (Comitato Etico Province di L'Aquila e Teramo, Approval number: reference protocol 52505/21, approval decree no. 1348/21). The study was conducted according to the principles of the Declaration of Helsinki.

Statistical Analysis

The risk of ASD in the offspring associated with the main exposure variable and the exposure covariates was studied by conditional logistic regression analysis. Adjustment for odds ratio (OR) estimates included all covariates determined to be a potential confounder by having a *p* value < .20 and changing the point estimate for technical classification by at least 10%. Missing covariates were accounted for by multiple imputation. The 95% confidence intervals (CI) for ORs were calculated by using Woolf's method. The effects of the polychotomous predictive variables were modelled by creating a set of dummy indicators [29-30].

We also carried out a sensitivity analysis, stratifying by offspring sex, to evaluate potential sex-specific associations. Moreover, we performed analyses restricted to level 1 (high-functioning) ASD cases – that is, the clinical subtype frequently exhibiting systemising abilities – to address heterogeneity of ASD.

Finally, we carried out a mediation analysis to investigate whether and the extent to which covariates found to be significantly associated with ASD risk mediate the relationship between parental education level and ASD in offspring. A regression-based approach under the counterfactual framework was used to perform the analysis, and the total effect of the exposure was decomposed to a controlled direct effect and an indirect effect. The controlled direct effect indicated the change in offspring ASD risk when the mediator was controlled at the reference level (e.g. without the presence of the mediator) and the exposure was changed from the reference to the index level. The indirect effect indicated the effect of the exposure that acted through the pre-specified mediator when the exposure was fixed to the reference. The proportion mediated estimate (PME), the measure of the proportion of the total effect of the exposure mediated by the intermediate variable on the log odds scale, was also calculated. If the PME is small, then the question is what are the other pathways through which the exposure affects the outcome, other than being a mediator. By contrast, if the PME is large, the total effect of the exposure on the outcome is through the mediator.

RESULTS

Risk of ASD in Offspring

Matching by age and sex ensured similar demographic characteristics of cases and controls. Table 1 describes the main characteristics of exposures for both groups and the results of logistic regression analysis.

Table 1. Conditional logistic regression: risk estimates (OR) of ASD for all exposures

Exposure variables	Exposure level	Cases	Controls	Unadjusted OR (95% confidence interval)	Adjusted OR (95% confidence interval)
Educational level of parents (highest between father or mother)	Low/middle	124	138	1.0*	1.0*
	Licence	345	381	1.01 (0.76-1.34)	0.95 (0.80-1.21)
	Degree (other)	73	79	1.03 (0.69-1.53)	0.94 (0.67-1.44)
	Degree (STEM)	116	60	**2.15 (1.45-3.19)**	**1.48 (1.03-2.55)**
	total	658	658		
Use of neurotropic drugs during mother's pregnancy	No	337	380	1.0*	1.0*
	Yes	39	52	0.85 (0.54-1.31)	0.90 (0.59-1.26)
	total	376°	432°		
Perinatal outcome (low birthweight and/or preterm birth)	No	321	357	1.0*	1.0*
	Yes	111	86	**1.43 (1.04-1.97)**	**1.29 (1.00-1.66)**
	total	432°	443°		
Socioeconomic status	Low	260	241	1.0*	1.0*
	Middle	281	291	0.90 (0.70-1.14)	0.92 (0.75-1.09)
	High	116	125	0.86 (0.63-1.17)	0.91 (0.68-1.06)
	total	658	658		
Familiarity	No	565	592	1.0*	1.0*
	Yes	92	66	**1.46 (1.04-2.04)**	**1.33 (1.00-1.88)**
	total	658	658		

*reference category
°totals are lower than the overall number of participants due to missing information

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Regarding the education level of parents, 189 out of 658 (28.7%) cases have at least one parent with a degree, compared with 139 out of 658 (21.1%) controls: this difference increases when considering only the STEM disciplines (116 out of 658 cases, 17.6%, vs 60 out of 658 controls, 9.1%), which comprise the majority of all degrees in parents of cases (116 out of 189, 61.4%) compared with the parents of controls (60 out of 139, 43.1%). After adjusting for all covariates, a parent with a STEM degree was significantly associated with risk of ASD in offspring, although the strength of association is low, as shown by interval estimates of the OR (1.43, 95% CI [1.03, 2.54]). Familiarity was associated significantly with the risk of ASD, although at the limit level (OR 1.33, 95% CI [1.00, 2.18]). Perinatal outcome (low birthweight and/or preterm birth) was associated significantly with risk of ASD at the limit level (OR 1.29, 95% CI [1.00, 1.66]). None of the other covariates were determinants of ASD risk in offspring.

Sensitivity Analysis

We did not find deviations from the overall analysis comparing the risk association in males and females, or when considering only level 1 ASD cases.

Mediation Analysis

Table 2 shows controlled direct and indirect effect ORs and PME. Familiarity was the strongest mediator on ASD risk in offspring (PME 28.0%). A weaker mediation effect was observed for perinatal outcomes (PME 11.5%).

Table 2 - Estimated direct and indirect effects of mediators on the association between main exposure and risk of autism spectrum disorder (ASD) in offspring. Only covariates significantly associated with ASD risk were considered in the model.

Main Exposure	Mediators	Controlled direct effect [a] OR (95% CI)	Indirect effect [b] OR (95% CI)	Proportion mediated estimate (PME)
Parent education Level (STEM)				
	Perinatal outcomes (low birthweight and/ or preterm birth)	1.35 (1.02-1.70)	1.06 (1.00-1.14)	11.5%
	Familiarity	1.51 (1.10-1.96)	1.10 (1.05-1.18)	28.0%

a. The controlled direct effect indicates the change in offspring ASD risk when the mediator was controlled at the reference level (e.g. without the presence of the mediating factor) and the exposure was changed from the reference to the index level.

b. The natural indirect effect indicates the effect of the exposure that acts through the mediator when the exposure was fixed to the reference.

DISCUSSION

We have addressed a topic of increasing interest in recent years – the risk of ASD in the offspring of parents with elevated systemising abilities – by using the achievement of a degree in STEM disciplines as a proxy characteristic. The idea that individuals with ASD might be predisposed to choosing and succeeding in STEM-related majors and jobs has gained a footing in the scientific literature [31-32]. We found that having a parent a STEM degree was significantly associated with ASD in offspring: 61.4% (116/189) of parents of individuals with ASD with a degree had chosen STEM disciplines. In addition, familiarity was associated significantly with ASD, as confirmed by the mediation analysis between parental education level and children with ASD. In this analysis, familiarity was the strongest mediator on ASD risk in offspring.

We also found that perinatal outcome (low birthweight and/or preterm birth) was associated significantly with ASD at the limit level. In the mediation analysis, however, this variable had a weaker mediation effect. These findings are in line with the literature, although the topic is still complex. Indeed, research has for some time now supported the link between autism and scientific disciplines [33]. Baron-Cohen et al. [34] reported the first-ever prospective study of a child born from adults with a formal diagnosis of Asperger syndrome (the child's parents were both scientists).

In the wake of these results, Dickerson et al. [24] showed that fathers of children with ASD were more than twice as likely to be engineers compared with fathers of children typical development. Furthermore, in a population-based study in San Francisco, Windham and collaborators [16] demonstrated that the risk of having a child with ASD was almost two times greater for mothers in highly technical occupations. However,

past research about parental occupation and ASD have yielded conflicting results. Specifically, Jarrold and collaborators [14] argued that previous research had not account for other systemising occupations in their analysis, showing that fathers of children in the same sample were also more likely to be accountants, scientists and physicians. Dickerson and collaborators [24] showed that after adjusting for demographic variables in their analysis, the higher likelihood of having children with ASD for parents who are engineers was no longer statistically significant. However, this increased likelihood of having a child with ASD remained significant for fathers employed in the fields of health care and accounting/financial analysis, even when accounting for demographic variables. Thus, to prevent possible biases due to socioeconomic abilities in seeking diagnosis and obtaining treatment for children with ASD, Dickenson et al. [24] also assessed the relationship between workers in an office or professional environment versus workers in manual jobs, as well as ≥ 16 years of education and having a child with ASD. They found no associations. They also found no significant relationship between paternal, maternal or joint technical occupation characteristics and ASD diagnosis. Windham et al. [16] also considered the socioeconomic status of parents; they found no significant association with paternal technical occupation and ASD diagnosis of offspring. On the contrary, they found a significant association between maternal occupations in highly technical fields and having offspring with ASD, with sustained significance in mothers in computer programming.

This research topic is still widely debated, but the unanimous result in the literature is that individuals with ASD are more likely than the general population and other groups of disabilities to gravitate towards STEM. Our manuscript goes in the direction of confirming the risk of occurrence of ASD in the offspring of parents with elevated systemising abilities. Van der Zee and Derksen [35] emphasised the power of systematisation in autism. The authors support that high systemising abilities are characteristic in autism. Systemising expands folk physics by understanding the laws or rules governing non-causal systems. Systematic thinking is correlated strongly with interests and predisposition to scientific studies. Systemising refers to involving or using a system. A system is anything that takes inputs and delivers outputs that the human brain can analyse, for example, technical systems (computer), abstract systems (mathematics), social systems (business) or organisable systems (library) [36-37]. Systemising allows the brain to predict that event x will occur with probability *p* [22]. Systemising is considered the most powerful way to predict change, because it involves the search for patterns. Several studies suggest that individuals with ASD have their systemising mechanism set at levels above those in the typical population. When one's systemising level is above average, this person could be considered to be immersed in the world of things rather than people. The reason for a higher systemising level in autism could be due to the idea that systemising increases the feeling of control [38]. In fact, we know that for people with autism it is more reassuring to control events and to avoid forms of the unexpected. Despite the published results and the scientific interest about this topic, we need further confirmation and other studies.

There are several limitations of our study.

First, the main limitation is the lack of important variables such as environmental or dietary exposures or genetic profiling. Although it is true that several studies have proposed a co-partnership between genetic and environmental factors as the aetiology of autism, these have never been specifically determined, but only highly suspected by indirect means. However, no responsible gene has ever been identified to date. None of genetic or 'growth ' hypotheses are taken into account as variables for this study.

Second, a highly schematic discipline may be a likely career choice for a person with ASD, precisely because of the forma mentis underlying the autistic spectrum: a predilection for repetitive patterns, difficulty in social interactions and difficulty in the act of "reading between the lines" (the individual with autism tends to stop at the primary and immediate meaning of sentences, without a more "metaphorical" reading). The individual would therefore do well to work in a STEM discipline, because it is systematic and easier to understand effectively. This implies that a not negligible proportion of parents is likely to have a profile overlapping the autistic spectrum and this may represent a source of relevant bias.

Third, the main exposure variable taken into account in the study is the cultural level of the parents, divided into low/medium, diploma, degree, STEM degree. This variable is heterogeneous between both parents and can lead to significant bias, i.e. the possibility of under- or over-diagnosis of the spectrum: the sensitivity to the topic by parents from different cultural backgrounds is not equal, even for simple knowledge of the existence of the disorder, especially if it is not severe.

Finally, there are inherent difficulties in the definition of ASD [39].

Last, we believe that a prospective study following STEM graduates over time and monitoring offspring would be optimal - although inefficient and expensive – if compared with a case-control study.

Future research should record the parental occupation at the time of the child's birth and at the time of clinical assessment in larger samples to investigate more possible confounders and effect modifiers. Moreover, more data about the professional training of parents (e.g. college majors and degrees earned) should be collected to further assess their inclination to choose highly structured career paths regardless of whether they eventually enter the workforce. These studies should contribute to enrich knowledge and eventually lead to advances in information on the aetiology of ASD.

Paper invited by Dr. Carlo La Vecchia, Editor-in-chief EBPH

CONTRIBUTORSHIP

Valenti M and Mazza M conceived, supervised and reviewed the study, finalized the manuscript and edited the paper. Attanasio M, Le Donne I and Bologna A collected and interpreted the data and gave substantial contribution in writing the paper; Masedu F and Tiberti S analysed the data, reviewed, critically revised and finalized the paper.

COMPETING INTERESTS

Authors have not competing interests.

FUNDING

The study was granted by public funds for research health programs in autism following the national Law no. 134 of 2015 (source: National Institute of Health and the Abruzzo Region of Italy Health System program VERGILIUS). A grant number is not available.

DATA SHARING

Data will be available upon reasonable request.

ETHICS APPROVAL

The study was approved by the Ethical Committee of the Local Health Agency (Comitato Etico Province di L'Aquila e Teramo, Approval number: reference protocol 52505/21, approval decree no. 1348/21).

REFERENCES

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. 2013.
2. Lai MC, Kassee C, Besney R, et al. Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *Lancet Psychiatry*. 2019;6:819-829. doi:10.1016/S2215-0366(19)30289-5
3. Vivanti G. Ask the Editor: What is the Most Appropriate Way to Talk About Individuals with a Diagnosis of Autism?. J Autism Dev Disord. 2020;50:691-693. doi:10.1007/s10803-019-04280-x
4. The Lancet Neurology. Neurology in the aftermath of Brexit. Lancet Neurol. 2016;15:995. doi:10.1016/S1474-4422(16)30164-8
5. Fombonne E, Morotti H, Mastel S, Keller K, Barnard RA, Hall T and O'Roak BJ. Autism questionnaire scores do not only rise because of autism. Dev Med Child Neurol; 2021;63:235-236.
6. Baio J, Wiggins L, Christensen DL, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveill Summ. 2018;67:1-23. doi:10.15585/mmwr.ss6706a1
7. ASDEU Consortium. Autism Spectrum Disorders in the European Union (ASDEU): final report: main results of the ASDEU project-28/08/2018. 2018. Available from: http://hdl.handle. net/10400.18/6188
8. Matson JL and Kozlowski AM. The increasing prevalence of autism spectrum disorders. Research in Autism Spectrum Disorders 2011;5:418-25.
9. Stubbs EG, Ritvo ER, Mason-Brothers A. Autism and shared parental HLA antigens. J Am Acad Child Psychiatry. 1985;24:182-185. doi:10.1016/s0002-7138(09)60445-3
10. Baharnoori M, Bhardwaj SK, Srivastava LK. Neonatal behavioral changes in rats with gestational exposure to lipopolysaccharide: a prenatal infection model for developmental neuropsychiatric disorders. Schizophr Bull. 2012;38:444-456. doi:10.1093/schbul/sbq098
11. Mazza M, Pino MC, Keller R, et al. Qualitative Differences in Attribution of Mental States to Other People in Autism and Schizophrenia: What are the Tools for Differential Diagnosis?. J Autism Dev Disord. 2022;52:1283-1298. doi:10.1007/s10803-021-05035-3
12. Wei X, Yu JW, Shattuck P, McCracken M, Blackorby J. Science, technology, engineering, and mathematics (STEM) participation among college students with an autism spectrum disorder. J Autism Dev Disord. 2013;43:1539-1546. doi:10.1007/s10803-012-1700-z
13. Baron-Cohen S. Are children with autism superior at folk physics?. New Dir Child Dev. 1997;75:45-54. doi:10.1002/cd.23219977504
14. Jarrold C and Routh DA Is there really a link between engineering and autism? A reply to Baron-Cohen et al. Autism 1997;1:101-9.
15. Wheelwright S, Baron-Cohen S. The link between autism and skills such as engineering, maths, physics and computing: a reply to Jarrold and Routh. Autism. 2001;5:223-227. doi:10.1177/1362361301005002010
16. Windham GC, Fessel K, Grether JK. Autism spectrum disorders in relation to parental occupation in technical fields. *Autism Res*. 2009;2:183-191. doi:10.1002/aur.84
17. Roelfsema MT, Hoekstra RA, Allison C, et al. Are autism spectrum conditions more prevalent in an information-technology region? A school-based study of three regions in the Netherlands. J Autism Dev Disord. 2012;42:734-739. doi:10.1007/s10803-011-1302-1

18. Baron-Cohen S, Wheelwright S, Spong A, Scahill V and Lawson J. Are intuitive physics and intuitive psychology independent? The journal of developmental and learning disorders. 2001;5:47-78.
19. Pring L, Ryder N, Crane L, Hermelin B. Creativity in savant artists with autism. Autism. 2012;16:45-57. doi:10.1177/1362361311403783
20. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord. 2001;31:5-17. doi:10.1023/a:1005653411471
21. Baron-Cohen S, Wheelwright S, Stott C, Bolton P and Goodyer I. Is there a link between engineering and autism? Autism. 1997;1:101-9.
22. Baron-Cohen S. The hyper-systemizing, assortative mating theory of autism. Prog Neuropsychopharmacol Biol Psychiatry. 2006;30:865-872. doi:10.1016/j.pnpbp.2006.01.010
23. Buchen L. Scientists and autism: When geeks meet. Nature. 2011;479:25-27. Published 2011 Nov 2. doi:10.1038/479025a
24. Dickerson AS, Pearson DA, Loveland KA, Rahbar MH, Filipek PA. Role of parental occupation in autism spectrum disorder diagnosis and severity. Res Autism Spectr Disord. 2014;8:997-1007. doi:10.1016/j.rasd.2014.05.007
25. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Washington, DC, American Psychiatric Assossation (4th ed.) 1994.
26. Lord C, Rutter M, Goode S, et al. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord. 1989;19:185-212. doi:10.1007/BF02211841
27. Lord C, Rutter M, DiLavore PC, Risi S, Gotham K and Bishop SL. Autism diagnostic observation schedule: ADOS-2. Los Angeles, CA: Western Psychological Services 2012
28. Valenti M, Vagnetti R, Masedu F, Pino MC, Rossi A, Scattoni ML, Mazza M, Di Giovanni C, Attanasio M, Filocamo A, Le Donne I and Siracusano M. Register-based cumulative prevalence of autism spectrum disorders during childhood and adolescence in central Italy. *Epidemiology, Biostatistics and Public Health*. 2019;16:e13226
29. Breslow NE and Day NE. International Agency for Research on Cancer. *Statistical methods in cancer research*. Lyon: International Agency for Research on Cancer 1980.
30. Schlesselman J. Case-Control Studies: Design, Conduct, Analysis. Oxford University Press, New York 1982
31. Stohlmann M, Moore TJ and Roehrig GH. Considerations for Teaching Integrated STEM Education. Journal of Pre-College Engineering Education Research (J-PEER), 2012; 2: Article 4.
32. Alston RJ and Hampton JL. Science and engineering as viable career choices for students with disabilities: A survey of parents and teachers. Rehabilitation Counseling Bulletin. 2020;43:158–164.
33. Baron-Cohen S. Does Autism Occur More Often in Families of Physicists, Engineers, and Mathematicians? Autism 1998; 2:296-301.
34. Baron-Cohen S, Scott F, Wheelwright S, et al. Can Asperger syndrome be diagnosed at 26 months old? A genetic high-risk single-case study. J Child Neurol. 2006;21:351-356. doi:10.1177/08830738060210040901
35. van der Zee E, Derksen JJL. The Power of Systemizing in Autism. Child Psychiatry Hum Dev. 2021;52:321-331. doi:10.1007/s10578-020-01014-4
36. Baron-Cohen S. Is Asperger syndrome necessarily viewed as a disability? *Focus on Autism and Other Developmental Disabilities*. 2002;17:186–91.
37. Baron-Cohen S. The extreme male brain theory of autism. Trends Cogn Sci. 2002;6:248-254. doi:10.1016/s1364-6613(02)01904-6
38. Baron-Cohen S. The essential difference: male and female brains and the truth about autism. Basic Books, Cambridge 2003
39. GBD 2019 Adolescent Mortality Collaborators. Global, regional, and national mortality among young people aged 10-24 years, 1950-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2021 Oct 30;398(10311):1593-1618.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Assessing the Use of GEE Methods for Analyzing Continuous Outcomes from Family Studies: Strong Heart Family Study

Xi Chen [(1)], Ying Zhang [(2)], Amanda M. Fretts [(3)], Tauqeer Ali [(4)], Jason G. Umans [(5)], Richard B. Devereux [(6)], Elisa T. Lee [(7)], Shelley A. Cole [(8)], Yan D. Zhao [(9)]

(1) MD, PhD, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX. ORCID 0000-0002-5359-8902
(2) MD, PhD, Center for American Indian Health Research, BSE, University of Oklahoma Health Sciences Center (OUHSC), Oklahoma City, OK. ORCID 0000-0002-5359-8902
(3) PhD, MPH, Cardiovascular Health Research Unit, Department of Epidemiology. University of Washington, Seattle, WA. ORCID 0000-0002-5358-2345
(4) MD, MPH, PhD, Center for American Indian Health Research, BSE, University of Oklahoma Health Sciences Center (OUHSC), Oklahoma City, OK. ORCID 0000-0002-9176-879X
(5) MD, PhD, MedStar Health Research Institute, MedStar Health Research Institute, Hyattsville, MD, and Georgetown-Howard University, Washington D.C. ORCID 0000-0002-2746-3350
(6) MD, Weill Cornell Medicine, New York, NY. ORCID 0000-0002-8542-4982
(7) PhD, Center for American Indian Health Research, BSE, University of Oklahoma Health Sciences Center (OUHSC), Oklahoma City, OK. ORCID 0000-0003-1826-3602
(8) PhD, Texas Biomedical Research Institute, San Antonio, TX. ORCID 0000-0002-2651-0127
(9) PhD, Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center (OUHSC), Oklahoma City, OK. ORCID 0000-0003-3448-0527

CORRESPONDING AUTHOR: Xi Chen, MD, PhD, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX. Email: xchen22@mdanderson.org.

SUMMARY

Background: Because of its convenience and robustness, the generalized estimating equations (GEE) method has been commonly used to fit marginal models of continuous outcomes in family studies. However, unbalanced family sizes and complex pedigree structures within each family may challenge the GEE method, which treats families as clusters with the same correlation structure. The appropriateness of using the GEE method to analyze continuous outcomes in family studies remains unclear. In this paper, we performed simulation studies to evaluate the performance of GEE in the analysis of family study data.
Methods: In simulation studies, we generated data from a linear mixed effects model with individual random effects. The random effects covariance matrix is specified as twice that of the pedigree matrix from the Strong Heart Family Study (SHFS) and other hypothetical pedigree structures. A Bayesian approach that utilizes the pedigree matrix was also conducted as a benchmark to compare with GEE methods with either independent or exchangeable correlation structures. Finally, analysis with a real data example was included.
Results: Our simulation results showed that GEE with independent correlation structure worked well for family data with continuous outcomes. Real data analysis revealed that all GEE and Bayesian approaches produced similar results.
Conclusion: GEE model performs well on continuous outcome in family studies, and it yields estimated coefficients similar to a Bayesian model, which takes genetic relationship into account. Overall, GEE is robust to misspecification of genetic relationships among family members.

Keywords: Bayesian; Generalized Estimating Equations; Kinship Matrix; Simulation; Strong Heart Family Study

DOI: 10.54103/2282-0930/20636
Accepted: 16th January 2023

INTRODUCTION

Generalized Estimating Equations (GEE) is a popular estimation approach used to fit marginal models on continuous outcomes in studies with repeated measurements or with clusters. Liang & Zeger first proposed the GEE method in *Biometrika* in 1986 [1]. By July 2021, their famous paper had achieved 19,086 citations. The popularity of GEE facilitates its incorporation in major statistical software, such as SAS, R, and STATA. A pivotal robustness property motivating widespread application of GEE is the high consistency and efficiency of the coefficient solution, no matter whether the working correlation structure is correctly specified.

Nevertheless, some previous studies indicated concerns about either the soundness of the theory or the proper use of GEE. For example, Crowder proposed that when the parameters used to calculate the working correlation matrix are uncertain in its definition, the asymptotic properties of the estimators can break down [2]. Mancl & Leroux revealed that the estimator yielded by the GEE model was fully efficient only for cluster-level covariates or covariates that are mean-balanced across clusters. In addition, the efficiency deceased as the variation in cluster sizes increased, and greater reductions occurred with higher between-cluster covariate variation [3]. Another study concluded that in GEE, misspecification of the correlation structure can be subject to a substantial loss in efficiency when covariates possess within-subject variability [4]. Furthermore, some critiques noted that GEE might not be the optimal model to use for data that are inherently unbalanced or for data with highly varied within-cluster correlation structures [5], although the systematic proof or simulations corresponding to this comment were not provided.

For family studies, such as the Strong Heart Family Study (SHFS) [6], data are correlated as a result of individuals being nested within each family. Depending on the scale of the study, the size of enrolled families can range from one to hundreds. In family studies, the kinship matrix is the statistical unit to store the information of relatedness among family members. Because of these wide-ranging family sizes, the kinship matrix can be complex and varies highly from one family to another. Such unbalanced family sizes and complex distribution of kinship matrix structures pose challenges in data analyses. Due to its convenience and popularity, GEE has been commonly used in data analysis in the Strong Heart Family Study [7-11]. However, without guidance from systematic simulation studies, it is unclear whether GEE is an appropriate approach with which to analyze family data.

When applying the GEE approach to family study data, there are a few concerns. First, to incorporate the kinship matrix defined among individuals, one random effect must be specified for each individual in the family study. Therefore, the total number of random effects is equal to the sample size. This contrasts with a typical GEE application in which random effects are defined at a cluster level. Moreover, GEE treated all families as clusters with an identical and simple correlation structure, which is far from the truth that the correlation structures among families are highly varied and can be very complex. Among the available GEE software packages, the correlation structures are predetermined and do not allow freedom in assigning distinct correlation structures across clusters.

In view of the aforementioned potential issues of applying the GEE method to analyze family data, we conducted simulation studies to evaluate the performance of the GEE method using a variety of simulation scenarios. Anticipating that the GEE method may not be appropriate to analyze family studies in certain scenarios, we also evaluated a Bayesian method proposed by Bae, Perls, & Sebastiani [12]. Their approach not only considers the within-cluster (a family) correlation by incorporating the kinship matrix in the model, but also avoids convergence issues due to the adoption of a singular value decomposition of the random effect covariance matrix. The Bayesian method was evaluated in the same way as the GEE method under the same simulation scenarios. At the end of this paper, we include analysis of SHFS data as an example of an application with which to compare GEE and Bayesian approaches.

The remainder of the paper is organized as follows. In Section 2, we described the statistical methods used in the study and briefly explained the derivation of the kinship matrix. In Section 3, we summarized the results from the simulation studies and analysis of the SHFS data. In Section 4, we highlighted the findings and discussed potential topics for future study.

METHODS

Conduct Simulation Using Linear Mixed Model (Conditional Model)

Linear mixed models (LMM) are commonly used to model continuous outcome variables obtained from correlated data. LMM include both fixed effects and random effects. In our study, to capture the kinship relationship among individuals with clusters, we specified random effects at individual levels.

Suppose we observe an outcome variable y in a sample with m families/clusters and a total sample size of n. Let n_i be the size for the ith family, $j=1,...,m$. The outcome for the jth individual in the ith family was generated by the model $y_{ij}=\mathbf{X_{ij}}\,\boldsymbol{\beta}+b_{ij}+\epsilon_{ij}$, where $\mathbf{X_{ij}}$ was the vector of covariates, $\boldsymbol{\beta}$ was the vector of fixed effects coefficients, b_{ij} was the individual-specific random effect that accounts for the additive polygenic effect, and ϵ_{ij} was the random error. For family i, we

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

stacked all of the random effects into a vector $\mathbf{b}_i=(b_{i1},...,b_{in_i})$, and we assumed $\mathbf{b}_i \sim N(\mathbf{0},\sigma_g^2 \mathbf{A}_i)$, where σ_g^2 was the unknown genetic variance and $\mathbf{A}_i$ was the known correlation matrix, which was 2 times the kinship matrix $\mathbf{K}_i$.

Kinship matrix is a matrix consisting of kinship coefficients between any pair of individuals. The kinship coefficient K_{rs} for any two individuals r,s is the probability that genes selected randomly from r and s from the same autosomal locus are inherited from a common ancestor. Because the kinship sampling is done with replacement, when $r=s$ that is for the same person, $K_{rs}=1/2$. Table 1 lists kinship coefficients for several common types of relative pairs [13].

Table 1. Kinship coefficients for several common types of relative pairs

Relationship	Parent - Offspring	Half Siblings	Full Siblings	First Cousins	Uncle - Nephew
Kinship coefficient	1/4	1/8	1/4	1/16	1/8

In our simulations, two independent variables, age (continuous) and gender (binary), were included. The fixed effects of age (β_1) and gender (β_2) were set as $\beta_1=0.08$ and $\beta_2=-0.5$. The value of intercept was set as $\beta_0=1$. The value of genetic variance was set as $\sigma_g^2=1$. The random error was generated from a standard normal distribution $\epsilon_i \sim N(0,1)$. The values of the kinship matrix and independent variables were provided separately in two sets of simulations described below. In each of the simulation scenarios, we conducted 1,000 runs.

The first set of simulations was conducted using information obtained from the SHFS, a family-based prospective cohort study of cardiovascular diseases (CVD) and its risk factors among American Indians from 12 tribal communities in central Arizona, southwest Oklahoma, and North and South Dakota [6]. In our project, we adopted the baseline data of the SHFS. A total of 91 families with 2,764 individuals were included. Family sizes ranged from 1 to 113, with a median of 31, Q1 16 and Q3 39, with 78% of family sizes less than 40. The values of age and sex from the SHFS were adopted as the vector of covariates X_i. The SHFS kinship coefficients, which were derived from participant interviews and other lab work, were directly used to build up the kinship matrix $K_i, i=1,...,91$.

The second set of simulations was performed based on hypothetical families with selected kinship structures. The second data scenario was constructed to mimic a different kind of family data, in which kinship coefficients were not provided directly. Instead, the kinship matrix was derived by an R package kinship2, which requires variables of individual ID, individual's father ID and mother ID, and family ID to process the algorithm [14]. As an example, Figure 1 shows the data frame for a nuclear family (a), the pedigree plot (b), and the kinship matrix (c) calculated by the kinship2 package.

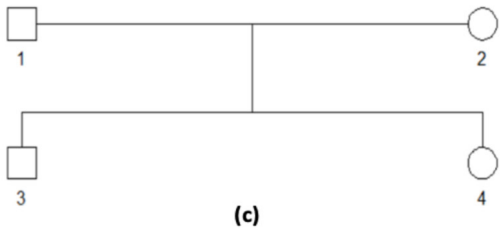
Figure 1. Data frame for a nuclear family (a), the pedigree plot (b), and the kinship matrix (c) calculated by the kinship2 package.

ID	Dad ID	Mom ID	Sex	Family ID
1	0	0	M	1
2	0	0	F	1
3	1	2	M	1
4	1	2	F	1

(a)

	1	2	3	4
1	0.50	0.00	0.25	0.25
2	0.00	0.50	0.25	0.25
3	0.25	0.25	0.50	0.25
4	0.25	0.25	0.25	0.50

(b)



(c)

Inspired by previous studies [12], we generated the corresponding variables of ID-series to create these family structures: (1) Singleton family: The family has only one member (same as independent data). (2) Nuclear family: The family structure is composed of a couple (father and mother) with two offspring. (3) Two-trios: This family structure is made up of first-, second-, and third-degree relatives, where two parent-offspring trios are related through a sibling pair in the parent generation. (4) Asymmetric family: This is an asymmetric and extended version of the second scenario, in which the first trio has only one offspring and the second trio has ten offspring.

In the second set of simulations, a family dataset with a total of 335 families and 1,020 individuals was generated. In each family, the gender of parents was defined as male as father and female as mother,

and the gender of children was randomly created by Bernoulli (0.5). The age of individuals was simulated by Uniform [a,b] for each generation of the family, with the boundaries of a and b set based on common logical order of parenthood, such that parents were older than the offspring and at least 25 years old.

For a linear outcome, the fixed effects coefficients in the conditional model and in the marginal model are equal mathematically. Since $y_{ij}=\mathbf{X_{ij}}\,\boldsymbol{\beta}+b_{ij}+\epsilon_{ij}$, and $\mathbf{b}_i \sim N(\mathbf{0},\sigma_g^2\,\mathbf{A}_i)$, $\epsilon_{ij}\sim N(0,1)$, then the expectation of the outcome of the conditional model is $\mathbf{E}(y_{ij})=E(\mathbf{X_i}\,\boldsymbol{\beta}+b_i+\epsilon_i)=\mathbf{X_i}\,\boldsymbol{\beta}$, which is the expectation of the outcome variable in the marginal model. Thus, the assumed values of fixed effects in the simulated conditional model can be directly used as the true values of the fixed effects to evaluate the marginal model.

Generalized Estimating Equations (GEE) (Marginal Model)

The generalized estimating equations (GEE) method is the most common method to fit marginal models for longitudinal/clustered data. The GEE method uses an iterative algorithm to estimate regression coefficients and variance-covariance matrix. Standard errors for the estimates of regression coefficients are computed using a robust sandwich estimator. The "working" correlation structure in the synthesis of variance-covariance described the pattern of correlations within clusters. The independent correlation structure and exchangeable correlation structures were used in our study, as they are the top choices of analysis performed on family studies. Independent correlation structure assumes that any two of the individuals are independent in a cluster. Exchangeable correlation structure assumes that any two of the individuals share the same correlation. A previous study recommended that exchangeable correlation structure should be used for observations within a cluster, but without logical ordering [15]. The R package *geeM* was used to implement the GEE [16].

Bae's Bayesian Approach

To compare with GEE, we compared a novel Bayesian approach, in which the kinship matrix was incorporated to account for the within-family correlation [12]. For the frequentist approach, models with random effects are used to capture the correlation among individuals in family studies. However, due to the large family sizes, the high dimensionality of the random effects vector makes it difficult to converge [13, 14]. Bae et al. proposed to incorporate the singular value decomposition (SVD) in the Bayesian modeling approach in family studies to improve the non-convergence issue. The SVD was applied on the large covariance matrix of the random effect to "break down" the high dimensionality. In particular, for each family, $\mathbf{A}_i$ is decomposed by SVD, $\mathbf{A}_i=\mathbf{U}_i\,\mathbf{S}_i\,\mathbf{U}_i'$, where $\mathbf{U}_i$ is the matrix of eigenvectors and $\mathbf{S}_i$ is the diagonal matrix of eigenvalues. Define $\mathbf{b}_i=\mathbf{G}_i\mathbf{u}_i$, where $\mathbf{G}_i=\mathbf{U}_i\,\mathbf{S}_i^{1/2}$ and $\mathbf{u}_i \sim N(\mathbf{0},\sigma_g^2\,\mathbf{I})$. We can show that $\mathbf{b}_i \sim N(\mathbf{0},\sigma_g^2\,\mathbf{A}_i)$. Therefore, the random effect $\mathbf{b}_i$ was replaced by $\mathbf{G}_i\mathbf{u}_i$. For example, the model function for continuous outcome can be rewritten as $y_i=\mathbf{X}_i\,\boldsymbol{\beta}+\mathbf{G}_i\mathbf{u}_i+\boldsymbol{\epsilon}_i$. Bae et al. used non-informative priors for the parameters and provided BUGS code. In our study, JAGS and the R package *rjags* were used in the Bayesian approach, since JAGS shared the same coding language with BUGS.

EVALUATION OF GEE AND BAYESIAN APPROACH

The GEE and the Bayesian approach were both performed on simulated data. Relative bias and coverage probability from these two approaches were used to assess the point and interval estimation. Relative bias was calculated as the absolute bias divided by the true values. The coverage probability was calculated by the proportions that the true value of coefficient lies within 95% confidence intervals (or credible intervals) of coefficients generated in each simulation.

REAL DATA EXAMPLE

To further compare the GEE and Bayesian approaches, we performed analysis on real data obtained from SHFS. Suppose we aimed to investigate the factors that are related to systolic blood pressure: age, sex, body mass index (BMI), diabetes status, smoking, and alcohol consumption. Missing data were less than 1%, so a complete case data analysis was conducted. GEE (with independence and exchangeable correlation structures) and Bayesian approaches were both performed using the same software packages as were used in the simulation studies. Point estimates, standard error (standard deviation for the Bayesian model), and 95% confidence intervals (95% credible intervals for the Bayesian model) were compared.

RESULTS

Table 2 summarizes the results for the first set of simulations, which integrated the data from SHFS. Overall, all three models showed good performance. The relative biases were all close to zero and the coverage probabilities were all close to 95%. The GEE model with independence correlation structure performed slightly better than did the other two models. Table 3 summarizes metrics to evaluate models for the

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

second set of simulations in which hypothetical family structures were used. The results were similar to the those from the first set of simulations. There were no discernable differences in relative biases and coverage probabilities among the three models, and the GEE model with independence correlation structure seemed slightly better than the other two models.

Table 2. Comparison relative bias and coverage probability between GEE and Bayesian model approaches in simulated data based on kinship coefficients from SHFS.

Model	GEE (Independent)	GEE (Exchangeable)	Bayesian Model[a]
Relative Bias			
Intercept	0.001	-0.004	-0.0004
Age	-0.0000007	0.0007	0.0005
Sex	-0.005	-0.003	-0.006
Coverage Probability			
Intercept	0.948	0.937	0.934
Age	0.943	0.941	0.954
Sex	0.952	0.941	0.937

[a]results with 1000 iterations, burn-in=100, chains=3, thin=2

Table 3. Comparison of bias, relative bias, and coverage probability between GEE and Bayesian models in simulated data based on kinship coefficients from a combination of singleton, nuclear, one-trio, two-trio, and three-trio families.

Model	GEE (Independent)	GEE (Exchangeable)	Bayesian Model[a]
Relative Bias			
Intercept	0.00009	-0.0007	0.0002
Age	0.0004	0.0005	0.0004
Sex	-0.003	-0.003	-0.004
Coverage Probability			
Intercept	0.95	0.953	0.945
Age	0.942	0.942	0.939
Sex	0.945	0.945	0.945

[a]results with 1000 iterations, burn-in=100, chains=3, thin=2

The descriptive summary of the variables in the real data example is presented in Table 4. Participants in the study were middle aged, with mean age of 41 years, and generally overweight, with mean BMI 31kg/m^2. The majority of the participants were female (60%). The percentages of factors of interest were: diabetes (41%), current smoking (36%), and current drinking (58%).

Table 4. Descriptive summary of variables selected from SHFS for the analysis of real data

Variable	Mean	SD	Missing
Age	40.9	17.27	None
BMI	31.26	7.48	23
SBP	123	16.87	14
	Count	**Percent**	**Missing**
Sex (Female)	1649	59.70%	None
Diabetes	1115	40.60%	18
Current smoking	997	36.2%	10
Current drinking	1588	57.7%	12

Table 5 summarizes the point estimates, standard error (standard deviation for the Bayesian model), and the 95% confidence interval (credible interval for the Bayesian model) of the coefficients in each model. In general, the metrics were similar among the three models. The point estimates of the coefficients were very close. The Bayesian model tended to give smaller standard deviations than did the GEE models because GEE used robust sandwich estimation for the covariance matrix, while the Bayesian model made explicit distributional assumptions. For the 95% CI, the majority of the intervals were similar among the three models. There were disagreements on two covariates: diabetes and current drinking. The 95% CI of the two covariates covered zero for the GEE with exchangeable correlation structure, but were above zero for the other two models. The results were consistent with the fact that the p-values for the two covariables were around .05 for all three models.

Table 5. Summary of point estimates and standard error of model coefficients for analyses of SHFS data

	GEE (Independent)	GEE (Exchangeable)	Bayesian Model[a]
Point Estimates			
Intercept	96.95	98.626	96.344
Age	0.41	0.41	0.416
Sex	-6.23	-6.328	-6.43
BMI	0.368	0.373	0.382
Diabetes	1.847	1.716	1.623
Current smoking	-0.113	-0.617	-0.334

Current drinking	1.487	2.267	2.204
Standard Error[b]			
Intercept	1.717	1.529	1.483
Age	0.023	0.023	0.018
Sex	0.681	0.666	0.549
BMI	0.05	0.044	0.039
Diabetes	0.89	0.886	0.637
Current smoking	0.719	0.698	0.589
Current drinking	0.786	0.734	0.61
95% CI[c]			
Intercept	(93.584, 100.316)	(93.629, 99.623)	(93.62, 99.31)
Age	(0.364, 0.456)	(0.364, 0.455)	(0.381, 0.449)
Sex	(-7.563, -4.895)	(-7.633, -5.024)	(-7.464, -5.406)
BMI	(0.27, 0.466)	(0.287, 0.459)	(0.31, 0.454)
Diabetes	(0.103, 3.59)	(-0.02, 3.452)	(0.382, 2.837)
Current smoking	(-1.523, 1.3)	(-1.985, 0.752)	(-1.525, 0.84)
Current drinking	(-0.053, 3.028)	(0.828, 3.705)	(0.99, 3.351)

[a] results with 2000 iterations, burn-in=100, chains=3, thin=5

[b] Standard deviation for the Bayesian Model

[c] Confidence Interval for GEE; Credible Interval for Bayesian Model

DISCUSSION

GEE serves as a handy tool for researchers to fit marginal models and make statistical inferences on clustered data, since this method can efficiently generate consistent estimates, regardless of the correct specification of within-cluster correlation structure. Data collected from a family study are clustered data. The unbalanced family sizes and the complex within-cluster relatedness may challenge the GEE performance. We evaluated the performance of GEE on simulated data with different types of family scenarios.

Our study is thus far the first to conduct systematic simulation studies to evaluate GEE in analysis of continuous outcomes in a family study. We simulated outcome data with covariates and kinship matrix from a real study, the SHFS, in which the kinship coefficients were generated based on meticulous interview and laboratory work. Furthermore, we included simulations with hypothetical family structures. We included a Bayesian model, which incorporated the kinship matrix in the modeling process, as a benchmark to compare with the GEE method. Results from the two sets of simulations indicated that both models work well, and there was no discernable difference between them. Moreover, the results of real data analyses revealed that the GEE and Bayesian models yielded similar estimates.

The performance of GEE on family data with continuous outcome was surprisingly good. When there is high correlation within responses, correct specification of correlation of responses potentially increases the efficiency. However, our integration of specific within-cluster correlation, the kinship matrix, did not bring much benefit in the Bayesian approach. When comparing the two GEE models, the independence correlation structure worked slightly better than the exchangeable correlation structure, which contradicts the fact that exchangeable correlation structure is recommended when there is no logical ordering for observations within a cluster [15]. It is possible that the simple structure improves the model fit efficiency. In conclusion, our results show that the GEE model performs well on continuous outcome in family studies, and it is robust to misspecification of genetic relationships among family members.

Our evaluation of GEE on family study focused on continuous outcomes, and our conclusion should not be simply applied to categorical or count outcomes. For continuous outcomes, the true values of regression coefficients in marginal models are equal to the values of fixed-effect regression coefficients in the conditional mixed effect models. This is because the linearity allows for the expectation of a continuous outcome to be calculated by the sum of expectation of each item in the model directly. However, categorical and count outcomes are typically modelled using a generalized linear model (marginal model) or a generalized linear mixed effects model (conditional model). Due to the nonlinearity of the link function, values of the regression coefficients in marginal models are no longer equal to those of in conditional models. Therefore, a direct comparison between the GEE approach and the Bayesian approach is not immediately available for categorical and count outcomes, and we leave this for future research.

ACKNOWLEDGEMENTS

The Strong Heart Study has been funded in whole or in part with federal funds from the National Heart, Lung, and Blood Institute, National Institute of Health, Department of Health and Human Services, under contract numbers 75N92019D00027, 75N92019D00028, 75N92019D00029, & 75N92019D00030. The study was previously supported by research grants: R01HL109315, R01HL109301, R01HL109284, R01HL109282,

and R01HL109319 and by cooperative agreements: U01HL41642, U01HL41652, U01HL41654, U01HL65520, and U01HL65521. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Indian Health Service (IHS). Statistical analysis was partially supported by National Institutes of Health, National Institute of General Medical Sciences [Grant 2U54GM104938, PI Judith James].

REFERENCES

1. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73(1):13-22.
2. Crowder M. On the use of a working correlation matrix in using generalised linear models for repeated measures. Biometrika. 1995;82(2):407-10.
3. Mancl LA, Leroux BG. Efficiency of regression estimates for clustered data. Biometrics. 1996:500-11.
4. Wang YG, Carey V. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. Biometrika. 2003;90(1):29-41.
5. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis: John Wiley & Sons; 2012.
6. North KE, Howard BV, Welty TK, Best LG, Lee ET, Yeh J, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. American journal of epidemiology. 2003;157(4):303-14.
7. Peng H, Zhu Y, Yeh F, Cole SA, Best LG, Lin J, et al. Impact of biological aging on arterial aging in American Indians: findings from the Strong Heart Family Study. Aging (Albany NY). 2016;8(8):1583.
8. Tinkelman NE, Spratlen MJ, Domingo-Relloso A, Tellez-Plaza M, Grau-Perez M, Francesconi KA, et al. Associations of maternal arsenic exposure with adult fasting glucose and insulin resistance in the Strong Heart Study and Strong Heart Family Study. Environment international. 2020;137:105531.
9. Jensen PN, Howard BV, Best LG, O'Leary M, Devereux RB, Cole SA, et al. Associations of diet soda and non-caloric artificial sweetener use with markers of glucose and insulin homeostasis and incident diabetes: the Strong Heart Family Study. European journal of clinical nutrition. 2020;74(2):322-7.
10. Grau-Perez M, Zhao J, Pierce B, Francesconi KA, Goessler W, Zhu Y, et al. Urinary metals and leukocyte telomere length in American Indian communities: the strong heart and the strong heart family study. Environmental Pollution. 2019;246:311-8.
11. Zhao Q, Zhu Y, Yeh F, Lin J, Lee ET, Cole SA, et al. Depressive symptoms are associated with leukocyte telomere length in American Indians: findings from the Strong Heart Family Study. Aging (Albany NY). 2016;8(11):2961.
12. 12. Bae HT, Perls TT, Sebastiani P. An efficient technique for Bayesian modeling of family data using the BUGS software. Front Genet. 2014;5:390.
13. Lange K. Mathematical and statistical methods for genetic analysis: Springer Science & Business Media; 2003.
14. Sinnwell JP, Therneau TM, Schaid DJ. The kinship2 R package for pedigree data. Hum Hered. 2014;78(2):91-3.
15. Horton NJ, Lipsitz SR. Review of software to fit generalized estimating equation regression models. The American Statistician. 1999;53(2):160-9.
16. McDaniel LS, Henderson NC, Rathouz PJ. Fast pure R implementation of GEE: application of the Matrix package. The R journal. 2013;5(1):181.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Phase III Failures for a Lack of Efficacy can be, in Significant Part, Recovered (Introducing Success Probability Estimation Quantitatively)

Daniele De Martini[(1)]

(1) Università degli Studi di Genova.

CORRESPONDING AUTHOR: Daniele De Martini, daniele.demartini@unige.it, ORCID 0000-0002-6937-5287.

SUMMARY

The rate of phase III trials failures is approximately 42-45%, and most of them are due to a lack of efficacy. Some of the failures for a lack of efficacy are expected, due to type I errors in phase II and type II errors in phase III. However, the rate of these failures is far from saturating the global failure rate due to a lack of efficacy.
In this work, the probability of unexpected failure for a lack of efficacy in phase III trials is estimated to be about 14%, with credibility interval (9%, 18%). These failures can be recovered through an adequate planning/empowering of phase II, and by adopting conservative estimation for the sample size of phase III. The software *SP4CT* (a free web application available at *www.sp4ct.com*) allows these computations. This 14% rate of unexpected failures gives that every year approximately 270,000 patients uselessly undergo a phase III trial with a large damage in individual ethics; moreover, the unavailability of many effective treatments is a considerable damage for collective ethics. The 14% of unexpected failures also produces more than $11bn of pure waste, and generates a much higher lack of revenue given by drugs' marketing.

Keywords: failure reasons; unexpected failures; ethics; waste; conservative sample size; software for sample size.

1. INTRODUCTION

The problem of the high rate of phase III trials that in recent years have failed, approximately 42-45% [1–4], has been widely discussed in our recent paper entltled "Empowering Phase II Clinical Trials to Reduce Phase III Failures" [5], where pros and cons of possible countermeasures have been presented.

In practice, phase III failures are due, for the most part, to a lack of efficacy (approximately 57-66% [1,6,7]). Moreover, given that other failures of this kind are labeled as failures for economic or commercial reasons or failures for safety (we will develop these concepts later), the actual failure rate due to a lack of efficacy is even higher than that reported above.

However, just some of these failures (for a lack of efficacy) are expected: in fact, expected failures of this kind are caused by type I errors committed in phase II and by type II errors of phase III. These statistical errors can not be completely avoided and, given usual settings and data available in the literature (e.g. [8]), they cause the failure of approximately 20-25% of phase III trials, corresponding to about 50-55% of failures.

Then, the global failure rate due to a lack of efficacy results much higher than that of failures due to a lack of efficacy that are expected (i.e. those due to statistical errors).

In the discussion presented in [5], the concept of *failures for a Lack of efficacy that are Not Expected* (viz. *LNE*), that is, failures due to a lack of efficacy minus those due to a lack of efficacy attributable to statistical errors, has been introduced. As far as this concept played a central role within the discussion of the abovementioned paper, the order of magnitude of the rate of *LNE* has been conservatively elicitated, and set at 10% of the trials run (i.e., approximately 22-24% of the failures). This datum supported the conclusion arguing the need of expanding phase II trials to increase phase III success rate.

Therefore, to focus on the rate of *LNE* is: scientifically, ethically, and economically relevant. Through this work, we aim at estimating the rate of *LNE*. Therefore, this paper can be considered a quantitative complement

DOI: 10.54103/2282-0930/20638
Accepted: 31th March 2023

to [5], and an addition to paragraphs 2 and 3 in the Introduction of the book *Success Probability Estimation with Applications to Clinical Trials* [9] (pp. XXIV-XXVI).

Finally, note that *LNE* failures can be recovered through an adequate planning: the conservative estimation of phase III sample size based on phase II data is a useful technique [9–12], and the software *SP4CT* is free web application (*www.sp4ct.com*) that allows these computations.

2. SETS AND PROBABILISTIC EVALUATION

2.1. Defining sets

Consider the following sets: F, representing the failures; L, failures for lack of efficacy; S, failures for safety reasons; C, failures for commercial or economic reasons; O, failures for other reasons; E, failures (lack of statistical significance) due to statistical errors, that is, type I errors in phase II and type II errors in phase III. Thus, we have: $L, S, C, O, E \subset F$. In particular $O = F \setminus L \cup S \cup C$.

In order to define unexpected failures for lack of efficacy, *LNE*, recall that it is given by failures for lack of efficacy minus failures for statistical errors belonging to L.

We remark that L is corrected by enlarging it, since the rates of L reported in the literature can be considered underestimated. This is due to the following facts:

a. failures reported as C are often function of L and S;
b. failures for both L and S are usually reported as S, since S is undoubtedly more serious. Consequently, some failures should be reallocated to L, according to the model adopted. To this aim, three models will be presented in the next section.

To conclude, denoting by L^c the corrected set of failures for a lack of efficacy, we have $LNE = L^c \setminus L^c \cap E$. We are interested in $P(LNE)$.

2.2. Calculating probabilities

$P(LNE)$ is given by $P(LNE|F) \times P(F)$, and $P(LNE|F) = P(L^c|F) - P(L^c \cap E|F)$. Then, $P(L^c|F)$ and $P(L^c \cap E|F)$ are computed according to different models (that follow in next section), where it is assumed that $P(F)$, $P(L|F)$, $P(S|F)$, and $P(C|F)$ are given. Moreover, $P(E)$ can be computed given the phase II false discovery rate $FDRII$, the phase III nominal power π, and the phase III type I error probability α. In particular, consider that the probability of running a phase III under the null coincides with the phase II false discovery rate (i.e. $P(H0) = FDRII$). Then, $P(E)$ can be obtained through the Total Probability theorem (i.e. $P(E)$ is the weighted sum of the probabilities of the expexted failures for effective and non effective treatments). Thus, we obtain: $P(E) = P(E|H0)\,P(H0) + P(E|H1)\,P(H1) = (1-\alpha/2)\times FDRII + (1-\pi)\times(1-FDRII)$.

3. MODELS

3.1. Model 1: reallocating a part of C

It is a fact that some failures labeled C are actually a function of safety and efficacy measures [13,14]. Thus, a subset of C, i.e. CR, has to be reallocated to either L or S. In practice, CR is divided into CRL and CRS, to be added to L and S, respectively. Then $CRL \cap CRS = \emptyset$, $CRL \cup CRS = CR \subset C$, and $L^c = L \cup CRL$.

Assume that $P(CR|C)$ is given, and consider $P(CRL|CR)$ and $P(CRS|CR)$ to be proportional to $P(L|F)$ and $P(S|F)$, respectively. In other words, the amount of CR reallocated to L is proportional to the amplitude of L. Given the assumptions above, we obtain that $P(CRL|F) = P(C|F)\,P(CR|C)\,P(L|F) \,/\, (P(L|F)+P(S|F))$, and consequently $P(L^c|F) = P(L|F) + P(CRL|F)$. Analogously, $P(S^c|F)$ is computed.

Now, the point is how to compute $P(L^c \cap E|F)$, which, under independence between E and the failure reasons (viz. L, S, C), would result $P(L^c|F) \times P(E|F)$. However, logic gives that $C \setminus CR$ does not contain parts of E (i.e. $C \setminus CR \cap E = \emptyset$), and therefore we can not exploit independence. Then, we assume E equally distributed over the sets that can contain it: L^c, S^c, O. Therefore we obtain $P(L^c \cap E|F) = (P(L^c|F) \times P(E|F)) \,/\, (P(L^c|F) + P(S^c|F) + P(O|F))$.

For example, if we set $P(F) = 0.43$, $P(L|F) = 0.6$, $P(S|F) = 0.15$, $P(C|F) = 0.2$, $P(CR|C) = 0.7$, $FDRII = 0.1$, $\pi = 0.85$, $\alpha = 0.05$ (the latter three giving $P(E) = 0.2325$), then we obtain $P(LNE) = 0.1301$.

3.2. Model 2: reallocating a part of S

This model develops point b) in section 2.1, arguing that failures for safety hide a relevant part of those for lack of efficacy, because the former are more serious than the latter. In particular, we assumed that if a treatment fails for both causes (i.e. S and L), then it is labeled S. This implies that a part of S has to be reallocated to F.

Consequently a certain amount of $P(S|F)$ has to be moved to $P(L|F)$. In detail $P(L^c|F) = P(L^c \cap S|F) + P(L^c \cap S^-|F) = P(L^c|S|F)P(S|F) + P(L^c|S^-|F)P(S^-|F)$. Note that $P(L^c \cap S^-|F)$ is in fact $P(L|F)$, giving $P(L^c|S^-|F) = P(L|F) \,/\, P(S^-|F)$. Now, assume that L^c failures have the same probability, in occurence with S failures or under different failures (this is absolutely reliable), giving $P(L^c|S) = P(L^c|S^-|F) = P(L \cap S^-|F)$. Finally, $P(L^c|F) = P(L|F)P(S|F)/P(S^-|F) + P(L|F)$.

To compute $P(L^c \cap E|F)$, independence cannot be advocated since the remaining part of S does not contain parts of E. Then, E is considered equally distributed over the sets that can contain it: L^c, C, O, and we obtain $P(L^c \cap E|F) = (P(L^c|F)P(E|F)) \,/\, (P(L^c|F) + P(C|F) + P(O|F))$.

3.3. Model 3: reallocating parts of *C* and *S*

In this final model, we mix the two reallocation criteria of the above paragraphs. Consequently, the probability of the corrected version of *L* is given by the original one plus that reallocated from *C* and that reallocated from *S*, according to the formulas given above. Thus, we obtain: $P(L^c|F) = P(L|F) + P(CRL|F) + P(L|F)P(S|F)/P(S^-|F)$.

To compute $P(L^c \cap E|F)$, once again independence cannot be applied. *E* is considered equally distributed over the sets that still can contain it, that now are just L^c and *O*. Then, we have $P(L^c \cap E|F) = (P(L^c|F)P(E|F))/(P(L^c|F) + P(O|F))$.

4. STATISTICAL COMPUTATION

4.1. Distribution assumptions

An exact computation of the estimate $P(LNE)$ is not possible, because: a) there is a certain variability among data concerning the estimates of failures due to different causes; b) the type II errors adopted for planning phase III trials and the rate of false positive findings in phase II are not precisely known.

Estimates of the probability of failure due to lack of efficacy (i.e. $P(L|F)$) found in the literature are 57% and 66% [1,6,7]. Assuming these two estimates as equally likely, and considering also likely the values within their range, $P(L|F)$ has been considered uniformly distributed in the range of the estimates, that is $P(L|F) \sim U(.57, .66)$.

Analogously, the probability of failure due to safety concerns ($P(S|F)$) and the probability of failure for economic or commercial reasons ($P(C|F)$) have been considered uniformly distributed in the range of their respective minimum and maximun estimates found in the literature [1,6,7], that is $P(S|F) \sim U(.09, .21)$ and $P(C|F) \sim U(.18, .22)$. Estimates of phase III failures go from 42% to 45% [1–4], so we set $P(F) \sim U(.42, .45)$.

Since failures for economic or commercial reasons are often based on utility functions depending on safety and efficacy measures [13,14], in practice a relevant part of them (i.e. *CR*) is reallocated to failures for safety or lack of efficacy (i.e. *S* or *L*).

The literature does not report estimates of the probability of *CR*. We discussed the problem with some authoritative colleagues, and we elicited $P(CR|C) \sim U(.5, .75)$.

The probability of launching a phase II trial when the treatment is ineffective has been set $P(FDRII) \sim U(.05, .14)$, because: a) in some phase II trials the launching rule is based on statistical significance with threshold 5% or higher; b) it has been estimated that the FDR in top medical literature is 14% [8], where phase II clinical trials represent an even higher class of experiments, so that *FDRII* has be assumed to be at most 14%. As it concerns type I and type II errors, since the power thresholds usually adopted in phase III trials are 80-90% we set $\pi \sim U(.8, .9)$, and $\alpha = 0.05$ according to the requirement of major national and transnational agencies.

Finally, the distributions introduced in this section, from that of $P(L|F)$ to that of π, have been considered independent.

4.2. Simulation

The aim of statistical computation is to obtain the distribution of $P(LNE)$. Then, a simulation has been performed, on the basis of distributional assumptions of section 4.1 and probabilistic calculation of section 3.2.

To approximate the distribution of $P(LNE)$ we started from simulating data from the distributions defined in section 4.1. In particular, 10^6 raw data have been generated from the joint distribution of $(P(L|F), P(S|F), P(C|F), P(F), P(CR|C), FDRII, \pi)$. If $P(L|F)+P(S|F)+P(C|F) > 1$, then these summands were rescaled to obtain sum 1 (e.g. $P(L|F)$ became $P(L|F)/(P(L|F)+P(S|F)+P(C|F))$). When $P(L|F)+P(S|F)+P(C|F) < 1$ there was no problem, since some (few) other failure causes are allowed in the model, according to related literature [1,6,7]. In practice, $P(L|F)$ and related probabilities have been rescaled 23% of the simulated raws (i.e. $P(P(L|F) + P(S|F) + P(C|F) > 1) \approx 0.23$); although this correction looks quite frequent and might look as a signal of model inadequacy, note that the extra probability generated by the simulation is quite small, since $P(P(L|F) + P(S|F) + P(C|F) > 1.05) \approx 0.03$.

Finally, $P(LNE)$ has been computed for each raw data, according to probabilistic calculation of section 3.2.

4.3. Results

Adopting Model 1 we obtained that the average of $P(LNE)$ was 13.4%, with the central 90% of data resulting in (8.8%, 18.0%) (this can be viewed as a credibility interval). Model 2 gave the average of $P(LNE)$ equal to 14.2%, with the central 90% of data resulting in (9.8%, 18.7%). The average of $P(LNE)$ given by Model 3 was 13.9%, with the central 90% of data resulting in (8.4%, 19.6%). Note that the latter results lie between those obtained with the two previous Models, since this latter approach is a mix of them, whereas the variability increases a bit.

5. SENSITIVITY ANALYSIS

In this section we modify some hypotheses or relax some assumptions made in the above sections, in

order to evaluate how the results on $P(LNE)$ change.

First, we introduce a different lower bound for $P(C|F)$. Although [1] did not report the rate of failures for commercial or economic reasons, an estimate of $P(C|F)$ can be obtained. Indeed, since the estimates of $P(L|F)$ and $P(S|F)$ where 0.66 and 0.21, respectively, it follows that $P(C|F) \leq 0.13$ (C may be not the only other cause of failure). Given that estimates from other sources were higher (i.e. 0.18, 0.22), we adopted 0.13 as the lower bound, and $P(C|F) \sim U(.13, .22)$. Under this different setting results vary just a bit.

Second, the distribution assumptions of section 4.1 are changed: Gaussian distributions has been used instead of the seven Uniform distributions previously adopted. In particular, $N(\mu, \sigma^2)$ substituted $U(a, b)$, where $\mu = (a + b)/2$ and $\sigma = (b - a)/4$ (i.e. $(\mu - 2\sigma, \mu + 2\sigma) \approx (a, b)$). With these settings, the distribution of $P(LNE)$ was a little tighter.

Table 1. Statistics of the distribution of P (LNE) under different models and settings.

	Mod1	Mod2	Mod3
Basic setting Mean	13.4	14.2	13.9
5th p-tile	8.8	9.8	8.4
95th p-tile	18.0	18.7	19.6
$P(C\|F) \sim U(.13, .22)$ Mean	13.4	14.2	14.0
5th p-tile	9.0	9.9	8.7
95th p-tile	17.9	19.1	19.6
Gaussian priors Mean	13.4	14.2	14.0
5th p-tile	9.6	10.4	9.4
95th p-tile	17.3	18.5	18.8

6. CONCLUSIONS

In a recent paper [5] we discussed the problem of the high rate of phase III failures, and presented pros and cons of possible countermeasures. In that paper, a central role was played by the rate of failure for a lack of efficacy not expected, *LNE*, which has been elicitated to be 10%. Here, we estimated this rate in more depth, through three different models.

Results were very close among the models: estimates of the rate of *LNE* were ap- proximately 14%, with 90% credibility interval approximately (9%, 18%). Thus, the elicitation has been confirmed by technical results. Moreover, a sensitivity analysis supported the estimates obtained.

Given that every year approximately 3,800 phase III trials are run with, on average, 500 patients each, the estimated 14% of *LNE* translates into an individual ethical loss [15] of 266,000 patients uselessly undergoing a phase III trial, annually. Moreover, the damage for collective ethics [15] is the unavailability of many effective treatments. Since the cost of each patient enrolled in a phase III is, on average, $42,000, the 14% of unexpected failures also produces more than $11bn of pure waste, and the loss of revenue given by drugs' marketing.

In fact, it is worth noting that *failures for a Lack of efficacy that are Not Expected* can be recovered through adequate planning: the above numbers argue for the need of empowering phase II trials, and for that of adopting conservative strategies for phase III sample size computation, in order to reduce phase III failures.

The software *SP4CT* allows conservative sample size estimation for phase III trials, and can help in determining phase II sample size on the basis of the overall probability of success of phase II *and* phase III. *SP4CT* is a free web application that can be run at *www.sp4ct.com*. Moreover, *SP4CT* performs profit computations [16] and is a useful tool for portfolio strategic planning.

The problem still open is whether enlarging phase II is worth it or not, given that resources are limited and that enlarging some phase II trials might imply that some other phase II would be not launched.

REFERENCES

1. Arrowsmith J. Phase III and submission failures: 2007-2010. Nature Reviews Drug Discovery 2011; 10(2): 1-1.
2. DiMasi JA, Reichert JM, Feldman L, Malins A. Clinical Approval Success Rates for Investigational Cancer Drugs. Clinical Pharmacology & Therapeutics 2013; 94: 329-335.
3. Thomas DW, Burns J, Audette J, Carroll A, Dow-Hygelund C, Hay M. Clinical Development Success Rates 2006-2015. BIO Industry Analysis, 2016.
4. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics 2018; 20 (2): 273-286.
5. De Martini D. Empowering Phase II Clinical Trials to Reduce Phase III Failures. Pharmaceutical Statistics, DOI:10.1002/pst.1980, 2019.
6. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. *JAMA* Internal Medicine 2016; 176(12): 1826-1833.
7. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nature Biotechnology 2014; 32(1): 40-51.

8. Jager LR, Leek JT. An estimate of the science-wise false discovery rate and application to the top medical literature. Biostatistics 2014; 15(1): 1-12
9. De Martini D. Success Probability Estimation with Applications to Clinical Trials. Wiley and Sons, Hoboken, 2013.
10. Chuang-Stein C. Sample Size and the Probability of a Successful Trial. Pharmaceutical Statistics 2006; 5: 305–309.
11. Wang SJ, Hung HMJ, O'Neill RT. Adapting the sample size planning of a phase III trial based on phase II data. Pharmaceutical Statistics 2006; 5: 85–97.
12. De Martini D. Adapting by calibration the sample size of a phase III trial on the basis of phase II data. Pharmaceutical Statistics 2011, 10(2): 89-95.
13. Patel N, Bolognese J, Chuang-Stein C, Hewitt D, Gammaitoni A, Pinheiro J. Designing PhII trials based on program-level considerations: a case study for neuropathic pain. Drug Information Journal 2012; 46: 439-454.
14. Antonijevic Z, Kimber M, Manner D, Burman C-F, Pinheiro J, Bergenheim K. Optimizing drug development programs: type 2 diabetes case study. Therapeutic Innovation and Regulatory Science 2013; 47: 363-374.
15. Lellouch J, Schwartz D. L'essai therapeutique: ethique individuelle ou ethique collective? Revue de l'Institut International de Statistique 1971; 39: 127-36.
16. De Martini D. Profit Evaluations When Adaptation by Design Is Applied. Therapeutic Innovation and Regulatory Science 2016; 50(2): 213-220.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Joint frailty model for recurrent events and a terminal event in the presence of cure fraction

Zahra Arab Borzu, PhD[(1)], Ahmad Reza Baghestani[(2)], Elaheh Talebi Ghane, PhD[(3)], Ali Akbar Khadem Maboudi, PhD[(4)], Ali Akhavan[(5)], Anahita Saeedi[(6)]

(1) Assistant Professor, Department of Biostatistics & Epidemiology, School of Public Health, Zahedan University of Medical Sciences, Zahedan, Iran
(2) Associate Professor of Biostatistics Department, Physiotherapy Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
(3) Assistant Professor of Biostatistics, Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran
(4) Associate Professor of Biostatistics Department, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
(5) Assistant Professor of Radiation Oncology, Isfahan University of Medical Sciences, Isfahan, Iran
(6) Department of Biostatistics, School of public Health & Health Sciences, University of Massachusetts, Amherst, MA, USA

CORRESPONDING AUTHOR: *Ahmad Reza Baghestani, Associate Professor of Biostatistics Department, Physiotherapy Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: baghestani.ar@gmail.com*

SUMMARY

The observations of repeated or recurrent events occur in many longitudinal studies. Furthermore, sometimes there may exist a terminal event such as death, which is strongly correlated with recurrent events. In many situations, a fraction of subjects who will never experience the event of interest during a long follow-up period is considered to be cured. In this article, we proposed a joint frailty model in the presence of cure fraction. The dependency is modeled by shared frailty that is contained in both the recurrent and terminal events hazard functions. It allows to estimate two separate sets of parameters on the recurrent, death, and cure model. We applied the maximum likelihood method under a piecewise constant hazard function for model fitting. The proposed model is evaluated by simulation studies and an application to a breast cancer data is provided.

Keywords: Breast cancer; Cure model; Joint model; Recurrent event; Terminal event.

INTRODUCTION

In many clinical or epidemiological studies, there are situation in which subjects are measured repeatedly over a fixed time. For instance, repeated episodes of hospitalization or experience asthma attacks, tumor recurrences. Many methodologies have been considered for the analysis of recurrent event data [1-6]. In many settings exists a terminal event such as death. Therefore, the terminal event may be strongly correlated with recurrent events. More explicitly, if the rate of the recurrent event is unusually low (high) in a subject, that subject is also subject to decreased (increased) rate of death. For example, recurrent asthma attacks during a follow-up, which can lead to death. In this case, the ordinary assumption of independent censoring can be violated and lead to biased estimates [4]. There are two major approaches to analyze recurrent events in the presence of a terminal event: The marginal models and the frailty models. Marginal models attend on the marginal rates of the recurrent and terminal events that can not specify the dependence between recurrent and terminal events [7-10]. Frailty models mostly apply a latent variable to account for the correlation between the recurrent and terminal events so that the two event processes

DOI: 10.54103/2282-0930/20639
Accepted: 14th May 2023

are independent given the frailty. For example, Huang and Wolf proposed a general joint frailty model to account for the informative censoring [11]. Liu et al., introduced a nonparametric maximum penalized likelihood method for estimating hazard function in a joint frailty model with right censoring and delayed entry [4]. Mazroui et al., suggested a joint frailty model to analyze recurrent events and death. They used two gamma-distributed frailties to allow for both the inter-recurrences dependence and the dependence between the recurrences and the survival times [12].

In recent years, the development of new drugs and treatment regimens has resulted in the significant number of patients in the population who are not susceptible to the event and live longer with diseases such as cancer; consequently, a cured fraction of the population exists. The use of standard survival models, for example, the Cox proportional hazard model for such data may be inappropriate since these models are based on the assumption that all the subjects experience the event with probability one so that the overall survivor function descends to zero, approximately. This assumption cannot be used in recent clinical trials and medical researches, because many subjects may never experience the event of interest if the follow-up period is sufficiently long. In such cases, cure models are widely applied. In this paper, we had a motivating example of patients with breast cancer (BC). A total of 357 patients received surgery to remove tumors. Two hundred and fifty-seven (72%) patients had no recurrence and death due to BC. We showed the Kaplan–Meier curve of disease-free survival (time to the first recurrence or death, whichever happened first) for patients with BC in Figure 1. There were very few events after 5 years of follow-up period, denoting the existence of a large proportion of cured patients. Ignoring the existence of "cured" patients leads to underestimation of the hazard and consequently overestimation of the overall survival of non-cured patients [13].

Many studies have been done on cure models [14-17]. In the context of recurrent event data, Rondeau et al., proposed a frailty model for the recurrent events in the presence of cure fraction [13]. Zhao et al., introduced a new model for recurrent with terminal events which can incorporated zero recurrence subjects [17]. Kim proposed a joint model for recurrent with a terminal event in the presence of cure fraction. The suggested model applied two types of deaths for the cure and susceptible groups, which would be regarded as competing risk with a missing cause [18]. Liu et al., proposed a joint frailty model for zero-inflated recurrent events in the presence of a terminal event. In that model, the frailty effect on recurrent and death rates is the same. In this article, we presented a joint frailty model in the presence of cure fraction for recurrent events and terminal event (death) by a shared gamma frailty in which the frailty can have different effects on recurrent events and death rates [19]. Thus, our model combined the features of the Liu et al. (2016) for patients who had no chance of experiencing the recurrent or death events from breast cancer, "cured patients", and the Liu et al. (2004) for the joint frailty analysis of recurrent and terminal events; the frailty effect on recurrent and death rates is the different. One advantage of our model is that it can estimate the effect of covariates on the recurrence and death times, and the cured probability, simultaneously. It can also reveal the degree of dependency between disease recurrence and death.

The remainder of the article is organized as follows. In Section 2, we introduced the joint frailty model in the presence of cure fraction and the estimation method. In Section 3, we presented the simulation studies and their results. In Section 4, we applied the proposed model to the analysis of a real dataset and a concluding discussion is presented in Section 5.

THE MODEL

Notations

We define notations and definitions that are used in the model. Let $T_{ij} = \min(X_{ij}, C_i, D_i)$ be the observed follow-up time so that X_{ij}, C_i and D_i correspond to the jth recurrent event time for ith subject $(i = 1,\ldots,N, j = 1,\ldots,n_i)$, the right-censoring time and the death time. Similarly, the terminal time denote by$T^* = \min(C_i, D_i)$. We consider a binary indicator for recurrent event as $\delta_{ij} = I(T_{ij} = X_{ij})$ so that if $n_i > 0$ then $\delta_{ij} = 1$ and a binary indicator for terminal event as $\Delta_i = I(T^*_i = D_i)$. S_{ij} indicate gap times (the time interval from previous to next recurrent event) so that $S_{ij} = T_{ij} - T_{i(j-1)}$ are independent with conditional on frailties and covariates. The observation for subject i is $O_i(t) \equiv \{S_{ij}, T^*_i, \delta_{ij}, \Delta_i\}$. Based on the theory of multivariate counting processes [4,14], $N^{D^*}_i(t) = I(D_i \le t)$ and $N^D_i(t) = I(X_i \le t, \Delta_i = 1)$ are the actual and the observed death indicator by time t, respectively. Similarly, we denote by $N^{R^*}_i(t)$ and $N^R_i(t) = N^{R^*}_i(\min(X_i, t))$ the actual and observed number of recurrent events, respectively. Let $Y_i(t) = 1_{(t \le T^*_i)}$ the at-risk indicator of subject i at time t. The observed and the actual number of recurrent events that occurs for ith in $[t, t+dt)$ is respectively $dN^{R^*}_i(t) = N^{R^*}_i((t+dt)^-) - N^{R^*}_i(t^-)$ and $N^R_i(t) = Y_i(t)dN^{R^*}_i(t)$. The process history of subject i up to time t, is represented as $H_{it} = \sigma\{Y(h), N^R_i(h), N^D_i(h), Z_i(h), \omega_i(h), 0 \le h \le t\}$. Where $Z_i(h)$ is the vector of covariates and $\omega_i(h)$ is shared frailty for subject i. Furthermore, recurrent event processes, death and censoring times assume to be continuous, therefore, in the simultaneous occurrences of recurrent and death events, we assume that death

happens first. The death event and the recurrent events intensity processes at t are $Y_i(t)h_i(t)dt = P(dN_i^R(t) = 1 | \mathcal{F}_{it^-})$ and $Y_i(t)\lambda_i(t)dt = P(dN_i^D(t) = 1 | \mathcal{F}_{it^-})$, respectively, where $h_i(t)dt = P(dN_i^{R^*}(t) = 1 | Z_i(t), \omega_i, D_i \geq t)$ and $\lambda_i(t)dt = P(dN_i^{D^*}(t) = 1 | Z_i(t), \omega_i, D_i \geq t)$.

Model for recurrent events and a terminal event

Following the model of Liu et al (2004), the joint model for the recurrent and terminal events given by:

$$\begin{cases} \lambda_i(t \mid \omega_i) = \omega_i \lambda_0(t) \exp(\beta' Z_i(t)) = \omega_i \lambda_i(t) \\ h_i(t \mid \omega_i) = \omega_i^{\alpha} h_0(t) \exp(\beta^* Z_i(t)) = \omega_i^{\alpha} h_i(t) \end{cases} \quad (2.1)$$

Where $\lambda_0(t)$ and $h_0(t)$ are baseline hazard functions for recurrent events and death respectively. The parameters β and β^* are regression coefficients vector associated with the covariate vector Z_i for recurrent event and death rates that could be different. The random effect ω_i takes into account the dependence between recurrent times and the death time. We assume ω_i have the gamma distribution with mean 1 and variance θ. When $\theta = 0$ implies that the random effects ω_i's are exactly 1, i.e., and heterogeneity in both recurrent and terminal events is only explained by Z_i. In the proposed model (2.1), the degree of dependence between recurrent and death times showed by α. The assumption is that $\alpha = 0$ that is $h_i(t)$ does not depend on ω_i, and terminal event (death) is non-informative for the recurrent events $\lambda_i(t)$, so that two rates $h_i(t)$ and $\lambda_i(t)$ are independent. When $\alpha = 1$, the effect frailty on recurrent events and death is the same. When $\alpha > 1$ the recurrent and death rates are positively correlated; higher frailty will result in earlier death. Inversely, $\alpha < 1$ demonstrates that subjects with higher frailty will be less likely to death.

Joint cure model for recurrent events and a terminal event

Let U be a binary variable that a subject will eventually $(U_i = 1)$ or never experience the event of interest $(U_i = 0)$. The survival function of T given by $S(t \mid z) = p\, S_u(t \mid z) + (1 - p)$. Where $S_u(t \mid z)$ is survival function for uncured subject and $p = Pr(U = 1)$.

In order to assess the relationship between Z_i and the probability of cure, a logit link function is used:

$$\log it(p_i) = \gamma^T Z_i. \quad (2.2)$$

Where γ is a parameter that is associated with the cure rate through covariate Z.

Following the model of Liu et al. (2016), the frailty proportional hazard model for recurrent events for subjects that are susceptible or not cured is:

$$\lambda_i(t \mid \omega_i, U_i = 1) = \lambda_0(t) \exp(\beta Z_i + \omega_i). \quad (2.3)$$

Similarly, hazard model for terminal event is:

$$h_i(t \mid \omega_i, U_i = 1) = h_0(t) \exp(\beta^* Z_i + \omega^{\alpha}{}_i). \quad (2.4)$$

Combining equations (2.2), (2.3) and (2.4) we have a joint model of the recurrent and terminal events with a cure fraction. In this case, a subject cured cannot experience any recurrent events, nor death due to the disease. Conditional likelihood for subject i th can be written as:

$$L(O_i \mid \omega_i) = L_{i1}^{I(n_i>0,\Delta_i=1)} L_{i2}^{I(n_i=0,\Delta_i=1)} L_{i3}^{I(n_i>0,\Delta_i=0)} L_{i4}^{I(n_i=0,\Delta_i=0)}$$

Where

L_{i1} is the likelihood of observing recurrent events $(n_i > 0)$ and death $(\Delta_i = 1)$,

L_{i2} is the likelihood of observing no recurrent events $(n_i = 0)$ and death $(\Delta_i = 1)$,

L_{i3} is the likelihood of observing recurrent events $(n_i > 0)$ and no death $(\Delta_i = 0)$,

L_{i4} is the likelihood of observing no recurrent events $(n_i = 0)$ and no death $(\Delta_i = 0)$.

That L_{i4} is cure on recurrent and terminal (death) events.

We can write:

$$L_{i1} = (1 - p_i) S_i^R(t_i \mid \omega_i, U_i = 1) \prod_{j=1}^{n_i} \lambda_i(t_{ij} \mid \omega_i, U_i = 1)$$
$$\times h_i(t^* \mid \omega_i, U_i = 1)^{\Delta_i} S^D{}_i(t^*{}_i \mid \omega_i, U_i = 1),$$

$$L_{i2} = (1 - p_i) S_i^R(t_i \mid \omega_i, U_i = 1) \times h_i(t_i^* \mid \omega_i, U_i = 1)^{\Delta_i}$$
$$\times S^D{}_i(t_i^* \mid \omega_i, U_i = 1),$$

$$L_{i3} = (1 - p_i) S_i^R(t_i \mid \omega_i, U_i = 1) \times \prod_{j=1}^{n_i} \lambda_i(t_{ij} \mid \omega_i, U_i = 1)$$
$$\times S^D{}_i(t_i^* \mid \omega_i, U_i = 1),$$

$$L_{i4} = p_i + (1 - p_i) S_i^R(t_i \mid \omega_i, U_i = 1) \times S^D{}_i(t_i^* \mid \omega_i, U_i = 1),$$

Where $S_i^R(t \mid \omega_i, U_i = 1)$ and $S_i^D(t^* \mid \omega_i, U_i = 1)$ are survival functions for the recurrent and death times for those not cured:

$$S_i^R(t \mid \omega_i, U_i = 1) = \exp(-\exp(\beta Z_i + \omega_i) \Lambda_0(t)),$$
$$S_i^D(t^* \mid \omega_i, U_i = 1) = \exp(-\exp(\beta^* Z_i + \omega^{\alpha}{}_i) H_0(t))$$

The $\Lambda_0(t)$ and $H_0(t)$ are cumulative baseline hazard function for the recurrent event and death respectively. The full loglikelihood is:

$$l(O) = \ln \prod_{i=1}^{N} \int_0^{\infty} L(O_i \mid \omega_i) \pi_{\theta}(\omega_i) d\omega_i \quad (2.5)$$

Where $\pi_{\theta}(\omega_i)$ is density function for frailty shared.

Estimation

To obtain the parameters estimation in proposed model, we utilize maximize likelihood technique to estimate different parameter $\Phi = (h_0(.), \lambda_0(.), \beta, \beta^*, \alpha, \theta, \gamma)$ due to the difficulty of solving the integral in the full log-likelihood (2.5), we used approach Gauss–Laguerre quadrature which is a numerical approximation of an integral using a weighted average of the integrand computed at M predetermined quadrature points $u_m (m = 1, 2, \ldots, M)$ over random effect ω_i. This the numerical approximation can be as such, $L(O) \approx \sum_{m=1}^{M} L(O_i \mid u_m)\pi_\theta(u_m)\nu_m$, with $u_m = \sqrt{2z_m}$ and $\nu_m = \sqrt{2\eta_m}\exp(z^2{}_m)$ Where η_m and z_m can be obtained from tables or algorithms, details of the procedure presented by [20,21]. Further, we apply a piecewise constant baseline hazard function for the estimation of baseline hazard functions in our estimation method. In the piecewise constant hazard function, we first divided the follow–up duration for recurrent events in to 5 intervals by 5th quantile (denoted by knots $Q_1^{\lambda}, Q_2^{\lambda}, \ldots, Q_5^{\lambda}$ and $Q_0^{\lambda} = 0$ or the smallest recurrent event time). We have:

$\tilde{\lambda}_0(t) = \lambda_{0k}$ for $Q^{\lambda}{}_{k-1} < t \le Q^{\lambda}{}_k$ where $k = 1, 2, \ldots, 5$ or

$$\lambda_0(t) = \sum_{k=1}^{5} \lambda_{0k} I(Q^{\lambda}{}_{k-1} < t \le Q^{\lambda}{}_k)$$

The cumulative baseline hazard function is

$$\Lambda_0(t) = \sum_{k=1}^{5} \lambda_{0k} \max(0, \min(Q^{\lambda}{}_k - Q^{\lambda}{}_{k-1}, t - Q^{\lambda}{}_{k-1}))$$

Following the similar procedure, we can create the piecewise constant baseline hazard function for death, denoted by $\tilde{h}_0(t)$ and $\tilde{H}_0(t)$ for cumulative baseline death hazard.

We use $\hat{H}^{-1}$ as a variance estimator, where H is the converged Hessian matrix of the log likelihood. Moreover, due to positively constraints on the parameter $(\theta > 0)$, we utilize the exponential transformation and their standard error calculated by the delta method [22].

After replacing cumulative baseline hazards in log-likelihood (2.5), the resulting log-likelihood can be maximized by the Gauss–Laguerre quadrature with implementation in R software.

SIMULATION

In this study, six hundred replicate datasets were generated, each with sample size (n=250, 500, 1000) to investigate the effect of increased sample size in parameters estimation. The simulation results of the parameters estimation are provided in Tables 1-4, which includes the Estimation parameter (Est), the empirical standard errors (SE), the mean square error (MSE), and the 95% empirical coverage probabilities (CP). The AIC mean and the number of propriety for the proposed and reduced models, which was the result of the minimum AIC value, were also reported, we considered the right-censored and utilized calendar time scale representation.

Generating Data

For each subject i, we generated binary explanatory variables $Z_i (i = 1, 2)$, from a Bernoulli distribution with probability 0.5. The random variables ω_i was generated from gamma distribution so that $\omega_i \sim gamma\left(\frac{1}{\theta}, \frac{1}{\theta}\right)$ with $\theta = 0.5$. A fixed right-censoring time was taken as $C_i = 6 + Unif(0, 6)$. We generated the gap times X_{ik} from $\lambda_i(s \mid \omega_i, U_i = 1) = \omega_i \lambda_0(s)\exp(\beta_1 Z_1 + \beta_1 Z_{2i})$ where $\lambda_0(t) = 0.65t^{0.25}$ and death time D_i generated from $h_i(t \mid \omega_i) = \omega^{\alpha}{}_i h_0(t)\exp(\beta_2 Z_{1i})$ where $h_0(t) = 0.4t^{0.25}$.

A death time D_i was generated from the hazard function $h_i(t \mid \omega_i)$.

If observed time was a death time $D_i \le C_i$ then $T_i^* = D$ and $\Delta_i = 1$.

If $D_i > C_i$ individual was censored then $T_i^* = C_i$ and $\Delta_i = 0$.

We used a logistic regression for probability of cure so that: $p_i = \frac{1}{1 + \exp(\gamma_0 + \gamma_1 Z_{i1})}$ and set $\alpha_0 = -0.5$ and $\alpha_1 = 1$.

We generated a random variable u_i from uniform distribution $[0, 1]$. The individual was cured (any recurrent nor death) if $u_i < p_i$ and individual was non-cured if $u_i \ge p_i$. The calendar times created from

$$T_{ij} = \min(C_i, D_i, \sum_{k=1}^{j} X_{ik})$$

If $T_{ij} < T_i^*$ then the observed time can be a recurrent event time and $\delta_{ij} = 1$. The data generating continues until $T_{ij} < T_i^*$.

If $T_{ij} \ge T_i^*$ individual was censored then $T_{ij} = T_i^*$ and $\delta_{ij} = 0$.

We set $\beta_1 = 1, \beta_2 = -0.5, \beta_1^* = 0.7, \alpha = 2, \gamma_0 = -0.5, \gamma_1 = 1$.

To compare the proposed model with two reduced models, we considered four different settings of $\alpha, \gamma_0, \gamma_1$ as following.

In setting I, we generated joint frailty model without cure fraction $(\gamma_0 = \gamma_1 = 10, \alpha = 2)$ Since $p_i = \frac{1}{1 + \exp(\gamma_0 + \gamma_1 Z_{i1})}$ for $(Z_{i1} = 0, 1)$, we had a mean

of cure percentage (p_i) close to zero. The estimates of parameters in the proposed model can be compared with the model of Liu et al., (2004).

In setting II, we generated joint frailty model in the presence of cure fraction ($\gamma_0 = -0.5\gamma_1 = 1, \alpha = 1$). For the situation, mean of cure percentage (p_i) close to 0.5 and the frailty effect on recurrent and terminal event rates is the same. The estimates of parameters in the proposed model can be compared with model of Liu et al., (2016).

In setting III, we generated joint frailty model with $\alpha > 1$, so that the recurrent rate and death are positively associated ($\gamma_0 = -0.5\gamma_1 = 1, \alpha = 2$). We can compare the estimates of parameters in the proposed model with the two reduced models (Liu et al., (2004) and Liu et al., (2016)).

In setting IV, we generated joint frailty model with $\alpha < 1$, so that the recurrent rate and death are negatively associated ($\gamma_0 = -0.5\gamma_1 = 1, \alpha = -2$). We can compare the estimates of parameters in the proposed model with the two reduced models (Liu et al., (2004) and Liu et al., (2016)).

Results of the simulation studies

The average numbers of deaths were 68% to 78%, the average numbers of recurrent events (among all 600 subjects) were 0.25 to 0.69 with a maximum fixed number of eight. The mean cure percentage was 50% in setting II and III.

In setting I, the mean cure percentage (p_i) was close to zero, so there was no cure fraction in datasets. In this case, both the joint frailty model (proposed model) and the reduced model (Liu et al, 2004) were equivalent. The mean of the estimates for γ_0 and γ_1 by the joint model are 9.937 and 10 respectively, which are very close to the true values.

It can be seen that the mean square errors and biases of parameters decreased with an increase in the sample size. In addition, AIC mean in proposed model was about four units more than the AIC mean in the reduced model, which was due to two extra parameters in the proposed model. Also, AIC percentage in the reduced model was lower than the proposed model in more 98.82% of cases. This indicates that even when the cure fraction does not exist, it is still valid to use the proposed model for data analysis.

In setting II, we had $\alpha = 1$ so there was a same correlation between recurrent and terminal event. The result showed that both the cure joint frailty model (proposed model) and the reduced model (II) were equivalent. The parameter estimates from these two models were virtually similar, with almost the same accuracy and precision. The mean of the estimates for α by the proposed model is 1.061, which is very close to the true value of $\alpha = 1$ in sample size 250. Thus, by increasing the sample size, α is underestimated.

We obtained a clear improvement in the estimates of parameters and mean square errors with increasing sample size. AIC mean in the proposed model was about one unit more than the AIC mean in the reduced model that by increasing the sample size, the difference raised to two. Furthermore, based on AIC percentage of all 600 replicate datasets, model (II) was preferred at least 81.8% times. This shows that when dependence between recurrent and terminal events is same, proposed model and model (II) are equivalent.

In setting III, we generated data from proposed model and set $\gamma_0 = -0.5, \gamma_1 = 1$. so that the mean of cure percentage was close to 0.5. We assumed $\alpha = 2$ which indicates significant positive dependence between recurrent and the death rates. In this case, proposed model is compared with two reduced models (Liu et al., (2004) and Liu et al., (2016)). The results of our model are summarized in the first panel in Table 3. The mean parameter estimates by new proposed joint frailty model were very close to their true values. There was a good agreement between the empirical and estimated standard errors of these parameter estimates, and the coverage probabilities were close to the nominal level of 95%. Moreover, the results show an underestimate for death risk and α which does not get better by increasing the sample size.

This can be due to the positivity constraint on the variance parameter. In comparison, we fit the model without the cure fraction. The results are reported in the third panel of Table 3. The results show that the absent of the cure fraction led to significant in biases and mean square errors in the estimate of parameters and very poor coverage probabilities. The estimate of the variance of the random effect in model without cure was much larger than that in our model (1.781 vs. 0.45). This shows that the new proposed cure joint frailty model can effectively capture the heterogeneity. Additionally, the lowest AIC mean and the high AIC percentage (98%) in the new proposed model suggests a better fit than two reduced models.

In setting IV, in order to assess a negative association between recurrent events and death rates we considered ($\alpha = -2$). Findings illustrate that the new proposed model offers very accurate parameter estimates and powerful coverage probabilities (Table 4).

APPLICATION-BREAST CANCER STUDY

Breast cancer (BC) is the most commonly diagnosed disease among females and includes 23% of total cancer cases with 14% risk of death. The cycle of this disease is usually determined by a response to initial treatment, followed by relapses. Moreover, relapse of breast cancer may increase the risk of death, which indicates an association between relapse and death.

In recent years, the improvement in treatment has led to 70-80% of patients being cured of BC. Common statistical models are not suitable for analyzing these data [23]. We applied the joint frailty model to analyze breast cancer (BC) with the new proposed model and two reduced models. Our real example is obtained from Shahid Ramezanzadeh Radiotherapy Center between April 2004 to March 2012; the patients were followed until April 2016. There were 357 females with BC included in the analysis. Among them, 77(21.6%) died, 69(19.3%) patients experienced recurrence of BC. The maximum number of recurrences for a patient was three. The numbers of patients with one, two and three recurrent events were 50(14%), 18(5%) and 1(0.3%), respectively. Two hundred and fifty-seven (72%) cases were cured, meaning that they experienced neither a recurrent event nor death due to BC. In this study, we considered four baseline covariates for each patient: Lymphovascular invasion (positive versus negative), age (50 years or older versus younger than 50 years), Lymph node status (positive versus negative) and tumor size (II, III versus I). Then we used the proposed joint model to analyze the effect of prognostic factors on recurrent and death times in the presence of cure fraction. For comparison, we also applied the joint frailty model without cure fraction and joint frailty model with the same frailty model in the presence of cure fraction, as introduced by Liu et al., in the years 2004 and 2016, respectively. In three models, the baseline hazard function is assumed to be piecewise constant for recurrent and terminal events, each with 5 intervals. The estimation results are shown in Table 5. We can see that the tumor size was significant in the cure model (P=0.013). The patients with larger tumor size were less likely to be cured. For illustration, hazard ratio of tumor size III and II were 0.77 and 0.48, respectively. Among those who were "not cured", tumor size was not significant. Patients with larger tumor sizes were more likely to experience recurrences. The hazard ratio of the patients with tumor size II and III were 1.012 and 1.008, respectively. In contrast, patients with larger tumor sizes had a lower morality rate than patients with tumor size I. Furthermore, we considered the same association between recurrent and terminal events leads to reduced model introduced by Liu et al., (2016), as shown in the second panel of Table 5. In this reduced model, sign and effect of variables were similar to those in our model except for the Lymph node status in death model, which showed that the patients with positive lymph node status were associated with a decreased risk of death (HR=0.704, P=0.304). We also fit another reduced model, which is a joint model without cure fraction introduced by Liu et al., (2004) as indicated in the third panel of Table 5. We noticed that the parameter estimates and their significance were different from those in the presence of cure fraction. The estimate of frailty variance without a cure fraction was more than that in the cure fraction model (variance estimate of θ increased from 0.904 to 1.288). This suggests that ignoring cure fraction leads to more heterogeneity for recurrent events in the reduced model. The positive values of α =1.357 to 1.8 show that the recurrence of disease and death rates were positively associated (P<0.001). The cured probability in our model and reduced model (II) was 77% and 85%, respectively. We obtained the cured probability in the data about 72%, indicating a more accurate estimate in our model. The Akaike information criterion (AIC) was also calculated, the AIC values indicated that the proposed model had a better fit than reduced models with the lowest value AIC=2062.

DISCUSSION AND CONCLUSION

In this paper, we introduced a joint frailty model in the presence of cure fraction. Our proposed model has two main advantages: on the one hand, the new joint frailty model can take into account a cure component. In this situation, the cured subject experience neither the recurrent events, nor death due to the diseases. On the other hand, our proposed model can evaluate the degree of dependence between recurrent and death times through the estimation parameter α. We have shown by simulation that using our joint frailty model in the presence of cure fraction led to unbiased regression coefficients, smaller mean square error, better coverage probabilities and less AIC in comparison with two reduced models. The simulation results show that in the presence of cure fraction, if $\alpha > 1$ and we falsely consider $\alpha = 1$, an underestimation of the recurrent and death rates occurs. In contrast, if $\alpha < 1$ and we falsely consider $\alpha = 1$, then recurrent and death rates is overestimated. The simulation results demonstrate that our proposed model is valid, even when there is the same dependence between recurrent and death times or there is non-cure fraction in the dataset. The proposed model was applied to a breast cancer dataset, and we showed that a positive association exists between recurrent and death rates. In this case, higher frailty implies an expected real death. In this article, we used gamma distribution for frailty. Other distributions can be used as well, e.g., Gaussian distribution (Liu et al., 2016). We have assumed piecewise for $\lambda_0(t)$ and $h_0(t)$. We can consider semi-parametric modeling (using spline function) for baseline hazard functions for recurrence and death, which provide more flexibility and reliable estimates of the cure fraction [12,13,24]. For the future works, our model can be more complex by considering longitudinal biomarkers and the joint with the recurrent model.

ACKNOWLEDGMENTS

The authors thank the reviewers and associate editor for their careful reading and valuable comments.

This study was approved by the Ethics Committees of the Shahid Beheshti University of Medical Science, Tehran, Iran (no: IR.SBMU.RETECH.REC.1400.697).

REFERENCES

1. Cai J, Schaubel DE. Marginal means/rates models for multiple type recurrent event data. Lifetime data analysis. 2004;10(2):121-38.
2. Cook RJ, Lawless JF. Analysis of repeated events. Statistical Methods in Medical Research. 2002;11(2):141-66.
3. Little RJ. Modeling the drop-out mechanism in repeated-measures studies. Journal of the american statistical association. 1995;90(431):1112-21.
4. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. Biometrics. 2004;60(3):747-56.
5. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika. 1981;68(2):373-9.
6. Sun J, Song PXK. Statistical analysis of repeated measurements with informative censoring times. Statistics in medicine. 2001;20(1):63-73.
7. Li QH, Lagakos SW. Use of the Wei–Lin–Weissfeld method for the analysis of a recurring and a terminating event. Statistics in medicine. 1997;16(8):925-40.
8. Lin D, Wei L, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2000;62(4):711-30.
9. Schaubel DE, Zhang M. Estimating treatment effects on the marginal recurrent event mean in the presence of a terminating event. Lifetime data analysis. 2010;16(4):451-77.
10. Wei L-J, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American statistical association. 1989;84(408):1065-73.
11. Huang X, Wolfe RA. A frailty model for informative censoring. Biometrics. 2002;58(3):510-20.
12. Mazroui Y, Mathoulin-Pelissier S, Soubeyran P, Rondeau V. General joint frailty model for recurrent event data with a dependent terminal event: application to follicular lymphoma data. Statistics in medicine. 2012;31(11-12):1162-76.
13. Rondeau V, Schaffner E, Corbiere F, Gonzalez JR, Mathoulin-Pelissier S. Cure frailty models for survival data: application to recurrences for breast cancer and to hospital readmissions for colorectal cancer. Statistical methods in medical research. 2013;22(3):243-60.
14. Andersen PK, Borgan O, Gill RD, Keiding N. Statistical models based on counting processes: Springer Science & Business Media; 2012.
15. Peng Y, Dear KB. A nonparametric mixture model for cure rate estimation. Biometrics. 2000;56(1):237-43.
16. Sy JP, Taylor JM. Estimation in a Cox proportional hazards cure model. Biometrics. 2000;56(1):227-36.
17. Zhao X, Wang J, Zhou X, Zhu Z. Recurrent events analysis in the presence of terminal event and zero-recurrence subjects. Communications in Statistics-Theory and Methods. 2015;44(4):710-25.
18. Kim YJ. Joint model for recurrent event data with a cured fraction and a terminal event. Biometrical Journal. 2020;62(1):24-33.
19. Liu L, Huang X, Yaroshinsky A, Cormier JN. Joint frailty models for zero-inflated recurrent events in the presence of a terminal event. Biometrics. 2016;72(1):204-14.
20. bramowitz M, Stegun IA. Handbook of mathematical functions: with formulas, graphs, and mathematical tables: Dover publications New York; 1972.
21. Golub GH, Welsch JH. Calculation of Gauss quadrature rules. Mathematics of computation. 1969;23(106):221-30.
22. Knight K. Mathematical Statistics. Chapman and Hall/CRC Texts in Statistical Science: New York. 2000.
23. Ewertz M, Jensen AB. Late effects of breast cancer treatment and potentials for rehabilitation. Acta Oncologica. 2011;50(2):187-93.
24. Zhangsheng Y, Liu L. A joint model of recurrent events and a terminal event with a nonparametric covariate function. Statistics in medicine. 2011;30(22):2683-95.

APPENDIX

Appendix A: construction of the log-likelihood for the proposed joint frailty model with calendar timescale.

In this appendix, we explain the structure of full likelihood $L(O \mid \omega_i) = \prod_i \int L(O_i \mid \omega_i) f(\omega_i) d\omega_i$, where $L_i(h_0(.), \lambda_0(.), \beta, \beta^*, \alpha, \theta) = L_i(O \mid \omega)$ for subject i and $(j = 1, 2, \ldots, n_i)$ such as, $\delta_{(i,n_i+1)} = 0$.

we calculated the conditional likelihood for the patients who experience the occurrence of disease and death $(n_i > 0, \Delta_i = 1)$, we have:

$$S_i^R(t_i \mid \omega_i) = \exp(-\omega_i \sum_{j=1}^{n_i+1} \int_{T_{i(j-1)}}^{T_{ij}} Y_i(t)\lambda_i(t)dt),$$

$$S_i^D(t_i^* \mid \omega_i) = \exp(-\omega_i^{\alpha} \int_0^{T_i^*} Y_i(t)h_i(t)dt)$$

$$L_{i1}(O \mid \omega_i) = (1-p_i)S_i^R(t_i \mid \omega_i, U_i = 1) \prod_{j=1}^{n_i} \lambda_i(t_{ij} \mid \omega_i, U_i = 1) \times h_i(t_i^* \mid \omega_i, U_i = 1)^{\Delta_i} S^D{}_i(t_i^* \mid \omega_i, U_i = 1)$$

$$= (1-p_i)\omega_i^{n_i} \prod_{j=1}^{n_i} \lambda_i(t_{ij})^{\delta_{ij}} \times \exp(-\omega_i \sum_{j=1}^{n_i+1} \int_{T_{i(j-1)}}^{T_{ij}} Y_i(t)\lambda_i(t)dt) \times (\omega_i h_i(t_i^*))^{\alpha\delta_i^*} \times \exp(-\omega_i^{\alpha} \int_0^{T_i^*} Y_i(t)h_i(t)dt)$$

2) We consider $\omega_i \sim G(\frac{1}{\theta}, \frac{1}{\theta})$ with probability density $f(\omega) = \frac{\omega^{(1/\theta - 1)} \exp(-\omega/\theta)}{\Gamma(1/\theta)\theta^{(\frac{1}{\theta})}}$

The contribution of marginal likelihood is obtained by integrating out the random effect (ω_i)

$$l_{i1}(O) = \frac{(1-p_i) \times \prod_{j=1}^{n_i+1} \lambda(t_{ij})^{\delta_{ij}} \times (h_i(T_i^*))^{\alpha\delta_i^*}}{\theta^{(\frac{1}{\theta})}\Gamma(\frac{1}{\theta})} \times \int_0^{\infty} \omega^{(\alpha\delta_i^* + n_i + \frac{1}{\theta} - 1)} \times \exp(-\omega_i \sum_{j=1}^{n_i+1} \int_{T_{i(j-1)}}^{T_{ij}} Y_i(t)\lambda_i(t)\,dt)$$

$$-\omega_i^{\alpha} \int_0^{T_i^*} Y_i(t)h_i(t)dt(-\frac{\omega_i}{\theta})d\omega_i$$

The contribution of marginal log-likelihood for individual i is:

$$l_{i1}(O) = \log(1-p_i) + \delta_{ij} \sum_{j=1}^{n_{i+1}} \log(\lambda_{ij}) + \alpha\delta_i^* \log(h_i(t_i^*)) - \frac{1}{\theta}\log(\theta) - \log(\Gamma(\frac{1}{\theta})) - \frac{1}{\theta}\log(\Gamma(\frac{1}{\theta}))$$

$$+\log\left[\int_0^{\infty} \omega^{(\alpha\delta_i^* + n_i + \frac{1}{\theta} - 1)} \times \exp(-\omega_i \sum_{j=1}^{n_i+1} \int_{T_{i(j-1)}}^{T_{ij}} Y_i(t)\lambda_i(t)\,dt) - \omega_i^{\alpha} \int_0^{T_i^*} Y_i(t)h_i(t)dt(-\frac{\omega_i}{\theta})d\omega_i\right]$$

In situation 2, we have subjects that do not experience the recurrent event $n_i = 0$ and observing death $\Delta_i = 1$. the contribution of marginal log-likelihood for individual i can write:

$$l_{i2}(O \mid \omega_i) = \log(1-p_i) + \alpha\delta_i^* \log(h_i(t_i^*)) - \theta\log(\frac{1}{\theta}) - \log(\Gamma(\frac{1}{\theta}))$$

$$+\log\left[\int_0^{\infty} \omega_i^{(\alpha\delta_i^* + \frac{1}{\theta} - 1)} \exp(-\omega_i \sum_{j=1}^{n_i+1} \int_{T_{i(j-1)}}^{T_{ij}} Y_i(t)\lambda_i(t)\,dt) - \omega_i^{\alpha} \int_0^{T_i^*} Y_i(t)h_i(t)\,dt - \frac{\omega_i}{\theta} d\omega_i\right]$$

In situation 3, we have subjects that experience the recurrent event $\delta_{ij} = 1$ but no observing death $\Delta_i = 0$. the

contribution of marginal log-likelihood for individual i can write:

$$l_{i3}(O \mid \omega_i) = \log(1-p_i) + \delta_{ij}\sum_{j=1}^{n_i}\log(\lambda(t_{ij})) - \frac{1}{\theta}\log(\theta) - \log(\Gamma(\frac{1}{\theta}))$$

$$+\log\left[\int_0^{\infty} -\omega_i \sum_{j=1}^{n_i+1}\int_{T_{i(j-1)}}^{T_{ij}} Y_i(t)\lambda_i(t)dt - \omega_i^{\alpha}\int_0^{T_i^*} Y_i(t)h_i(t)dt - \frac{\omega_i}{\theta}d\omega_i\right]$$

In situation 4, we have subjects that experience neither recurrence nor death from the disease

$$l_{i4}(O \mid \omega_i) = \log\left\{\frac{[p_i + (1-p_i)]\times\int_0^{\infty}\omega_i^{(\frac{1}{\theta}-1)}\exp(-\omega_i\sum_{j=1}^{n_i+1}\int_{T_{i(j-1)}}^{T_{ij}} Y_i(t)\lambda_i(t)dt - \omega_i^{\alpha}\int_0^{T_i^*} Y_i(t)h_i(t)dt - \frac{\omega_i}{\theta}d\omega_i}{\theta^{\frac{1}{\theta}}\Gamma(\frac{1}{\theta})}\right\}$$

We can obtain full log likelihood by sum of the four marginal contribution of log-likelihood for subject i as follows:

$$l(O) = \sum_{i=1}^{4} l_i(O)$$

We can obtained the log-likelihood for gap times with replace T_{ij} by S_{ij} and $\int_{T_{i(j-1)}}^{T_{ij}}$ by $\int_0^{S_{ij}}$ in expression of log-likelihood.

Table 1. *Simulation Results for a generated joint frailty model with different frailty effect in absent of cure fraction*

Sample size	Parameter	Proposed model					Model by Liu et al. (2016)				
		Est	SE emp	SE $(\sqrt{\hat{H}^{-1}})$	MSE	CP	Est	SE emp	SE $(\sqrt{\hat{H}^{-1}})$	MSE	CP
N=250	$\beta_1 = 1$	0.964	0.152	0.155	0.024	0.934	0.969	0.15	0.155	0.023	0.938
	$\beta_2 = -0.5$	-0.507	0.131	0.131	0.017	0.941	-0.506	0.131	0.131	0.017	0.939
	$\beta_1^* = 0.7$	0.617	0.239	0.246	0.064	0.941	0.622	0.239	0.246	0.063	0.941
	$\theta = 0.5$	0.46	0.105	0.101	0.013	0.944	0.463	0.103	0.101	0.012	0.943
	$\alpha = 2$	1.833	0.452	0.457	0.232	0.958	1.826	0.437	0.453	0.222	0.943
	$\gamma_0 = 10$	9.826	1.264	44.079	1.628	0.963	-	-	-	-	-
	$\gamma_1 = 10$	10	0	32424.04	0	0.98	-	-	-	-	-
	mean_AIC			1690.986					1687.116		
	Percent_AIC			1.52%					98.48%		
N=500	$\beta_1 = 1$	0.976	0.103	0.109	0.011	0.946	0.978	0.103	0.109	0.011	0.942
	$\beta_2 = -0.5$	-0.507	0.093	0.092	0.009	0.949	-0.507	0.093	0.092	0.009	0.949
	$\beta_1^* = 0.7$	0.614	0.166	0.171	0.035	0.912	0.616	0.165	0.17	0.034	0.915
	$\theta = 0.5$	0.465	0.074	0.072	0.007	0.921	0.467	0.074	0.071	0.007	0.927
	$\alpha = 2$	1.767	0.276	0.307	0.13	0.874	1.765	0.275	0.305	0.131	0.87
	$\gamma_0 = 10$	9.872	0.965	37.245	0.947	0.978	-	-	-	-	-
	$\gamma_1 = 10$	10	0	14439.45	0	0.987	-	-	-	-	-
	mean_AIC			2135.385					3357.693		
	Percent_AIC			15.5%					98.48%		
N=1000	$\beta_1 = 1$	0.978	0.078	0.077	0.006	0.941	0.979	0.077	0.077	0.006	0.943
	$\beta_2 = -0.5$	-0.503	0.064	0.065	0.004	0.944	-0.503	0.064	0.065	0.004	0.944
	$\beta_1^* = 0.7$	0.618	0.12	0.12	0.021	0.887	0.619	0.12	0.12	0.021	0.889
	$\theta = 0.5$	0.467	0.052	0.051	0.004	0.896	0.468	0.052	0.051	0.004	0.899
	$\alpha = 2$	1.732	0.199	0.212	0.111	0.712	1.731	0.198	0.211	0.111	0.705
	$\gamma_0 = 10$	9.937	0.858	28.489	0.74	0.975	-	-	-	-	-
	$\gamma_1 = 10$	10	0.001	9335.436	0	0.992	-	-	-	-	-
	mean_AIC			6692.079					6688.152		
	Percent_AIC			1.18%					98.82%		

AIC, Akaike information criterion; CP, coverage probability; MSE, mean square error; SE, standard error

Table 2. *Simulation Results for a generated joint frailty model with same frailty in presence of cure fraction*

Sample size	Parameter	Proposed model					Model by Liu et al. (2016)				
		Est	SE emp	SE $(\sqrt{\hat{H^{-1}}})$	MSE	CP	Est	SE emp	SE $(\sqrt{\hat{H^{-1}}})$	MSE	CP
N=250	$\beta_1 = 1$	0.981	0.259	0.253	0.068	0.945	0.983	0.259	0.252	0.068	0.948
	$\beta_2 = -0.5$	-0.511	0.207	0.201	0.043	0.955	-0.512	0.207	0.200	0.043	0.953
	$\beta_1^* = 0.7$	0.687	0.318	0.297	0.101	0.937	0.675	0.308	0.281	0.095	0.943
	$\theta = 0.5$	0.462	0.174	0.169	0.032	0.955	0.452	0.164	0.157	0.029	0.957
	$\alpha = 1$	1.061	0.431	0.429	0.189	0.948	-	-	-	-	-
	$\gamma_0 = -0.5$	-0.501	0.219	0.208	0.048	0.942	-0.501	0.219	0.208	0.048	0.943
	$\gamma_1 = 1$	1.018	0.292	0.281	0.086	0.953	1.019	0.293	0.281	0.086	0.953
	mean_AIC			1078.811					1077.80		
	Percent_AIC			15.17%					84.83%		
N=500	$\beta_1 = 1$	0.98	0.182	0.178	0.034	0.945	0.981	0.182	0.177	0.034	0.942
	$\beta_2 = -0.5$	-0.517	0.145	0.142	0.021	0.953	-0.517	0.144	0.14	0.021	0.952
	$\beta_1^* = 0.7$	0.675	0.209	0.203	0.044	0.947	0.678	0.206	0.199	0.043	0.948
	$\theta = 0.5$	0.477	0.123	0.123	0.016	0.957	0.462	0.112	0.112	0.014	0.948
	$\alpha = 1$	0.969	0.244	0.259	0.061	0.945	-	-	-	-	-
	$\gamma_0 = -0.5$	-0.505	0.151	0.147	0.023	0.96	-0.506	0.15	0.146	0.023	0.96
	$\gamma_1 = 1$	1.015	0.203	0.198	0.042	0.947	1.015	0.203	0.198	0.042	0.947
	mean_AIC			2135.385					2134.354		
	Percent_AIC			15.5%					85.5%		
N=1000	$\beta_1 = 1$	0.973	0.125	0.126	0.016	0.942	0.974	0.125	0.125	0.016	0.945
	$\beta_2 = -0.5$	-0.508	0.101	0.101	0.010	0.947	-0.508	0.101	0.099	0.010	0.947
	$\beta_1^* = 0.7$	0.657	0.142	0.141	0.022	0.942	0.668	0.142	0.140	0.021	0.935
	$\theta = 0.5$	0.485	0.091	0.088	0.008	0.948	0.468	0.081	0.08	0.008	0.932
	$\alpha = 1$	0.93	0.165	0.172	0.032	0.922	-	-	-	-	-
	$\gamma_0 = -0.5$	-0.503	0.109	0.103	0.012	0.953	-0.505	0.109	0.103	0.012	0.952
	$\gamma_1 = 1$	1.008	0.149	0.139	0.022	0.947	1.009	0.148	0.139	0.022	0.947
	mean_AIC			4255.608					4254.756		
	Percent_AIC			18.2%					81.8%		

AIC, Akaike information criterion; CP, coverage probability; MSE, mean square error; SE, standard error

Table 3. *Simulation Results for a generated joint frailty model with different frailty* ($\alpha > 0$) *on recurrent and death rate in presence of cure fraction*

Sample size	Parameter	Proposed model										Liu et al(2004)				
		Est	SE emp	SE $(\sqrt{\hat{H}^{-1}})$	MSE	CP	Est	SE emp	SE $(\sqrt{\hat{H}^{-1}})$	MSE	CP	Est	SE emp	SE $(\sqrt{\hat{H}^{-1}})$	MSE	CP
N=250	$\beta_1 = 1$	0.98	0.263	0.262	0.069	0.95	0.975	0.262	0.271	0.069	0.955	1.724	0.303	0.261	0.616	0.34
	$\beta_2 = -0.5$	-0.515	0.187	0.190	0.035	0.938	-0.518	0.189	0.210	0.036	0.937	-0.534	0.229	0.200	0.054	0.95
	$\beta_1^* = 0.7$	0.656	0.419	0.407	0.177	0.943	0.501	0.314	0.299	0.138	0.905	1.399	0.389	0.385	0.640	0.582
	$\theta = 0.5$	0.442	0.181	0.178	0.036	0.943	0.495	0.217	0.191	0.047	0.958	1.765	0.086	0.187	1.608	0.000
	$\alpha = 2$	2.010	0.726	0.768	0.527	0.955	-	-	-	-	-	1.426	0.245	0.301	0.390	0.300
	$\gamma_0 = -0.5$	-0.501	0.228	0.219	0.052	0.947	-0.493	0.232	0.222	0.054	0.948	-	-	-	-	-
	$\gamma_1 = 1$	1.016	0.296	0.290	0.088	0.947	1.019	0.298	0.292	0.089	0.948	-	-	-	-	-
	mean_AIC			1121.014										1190.82		
	Percent_AIC			63.33%										0.00%		
N=500	$\beta_1 = 1$	0.976	0.184	0.185	0.035	0.958	0.970	0.183	0.191	0.034	0.96	1.719	0.218	0.184	0.565	0.080
	$\beta_2 = -0.5$	-0.517	0.135	0.134	0.019	0.953	-0.520	0.136	0.147	0.019	0.950	-0.532	0.164	0.140	0.028	0.947
	$\beta_1^* = 0.7$	0.636	0.279	0.275	0.082	0.940	0.503	0.215	0.211	0.085	0.857	1.380	0.276	0.264	0.539	0.312
	$\theta = 0.5$	0.451	0.132	0.130	0.02	0.943	0.501	0.142	0.136	0.02	0.953	1.775	0.061	0.132	1.630	0.000
	$\alpha = 2$	1.836	0.421	0.474	0.204	0.960	-	-	-	-	-	1.383	0.164	0.201	0.408	0.053
	$\gamma_0 = -0.5$	-0.510	0.150	0.154	0.023	0.952	-0.504	0.152	0.155	0.023	0.955	-	-	-	-	-
	$\gamma_1 = 1$	1.019	0.204	0.204	0.042	0.935	1.021	0.205	0.205	0.042	0.932	-	-	-	-	-
	mean_AIC			2218.537										2360.208		
	Percent_AIC			81%										0.00%		
N=1000	$\beta_1 = 1$	0.973	0.133	0.130	0.018	0.957	0.966	0.132	0.134	0.019	0.958	1.711	0.159	0.130	0.531	0.005
	$\beta_2 = -0.5$	-0.512	0.093	0.094	0.009	0.957	-0.515	0.094	0.103	0.009	0.957	-0.526	0.111	0.099	0.013	0.940
	$\beta_1^* = 0.7$	0.617	0.188	0.190	0.042	0.928	0.495	0.149	0.148	0.064	0.718	1.354	0.189	0.183	0.463	0.060
	$\theta = 0.5$	0.450	0.093	0.093	0.011	0.927	0.501	0.097	0.097	0.009	0.947	1.781	0.044	0.093	1.642	0.000
	$\alpha = 2$	1.783	0.287	0.319	0.129	0.870	-	-	-	-	-	1.366	0.113	0.138	0.415	0.003
	$\gamma_0 = -0.5$	-0.508	0.105	0.108	0.011	0.948	-0.502	0.106	0.109	0.011	0.948	-	-	-	-	-
	$\gamma_1 = 1$	1.009	0.142	0.143	0.020	0.947	1.012	0.143	0.144	0.021	0.95	-	-	-	-	-
	mean_AIC			4419.620					4427.928					4705.045		
	Percent_AIC			98.16%					1.84%					0.00%		

AIC, Akaike information criterion; CP, coverage probability; MSE, mean square error; SE, standard error

Table 4. *Simulation Results for a generated joint frailty model with different frailty on recurrent and death rate* $(\alpha < 0)$ *in presence of cure fraction*

Sample size	Parameter	Proposed model					Liu et al. (2016)					Liu et al. (2004)				
		Est	SE emp	SE $(\sqrt{\hat{H}^{-1}})$	MSE	CP	Est	SE emp	SE $(\sqrt{\hat{H}^{-1}})$	MSE	CP	Est	SE emp	SE $(\sqrt{\hat{H}^{-1}})$	MSE	CP
N=250	$\beta_1 = 1$	0.998	0.199	0.198	0.04	0.943	1.094	0.188	0.152	0.044	0.921	1.705	0.315	0.251	0.597	0.385
	$\beta_2 = -0.5$	-0.5	0.157	0.166	0.025	0.942	-0.492	0.174	0.136	0.03	0.946	-0.469	0.276	0.223	0.077	0.943
	$\beta_1^* = 0.7$	0.649	0.348	0.342	0.124	0.948	0.401	0.22	0.212	0.138	0.735	0.934	0.227	0.233	0.107	0.831
	$\theta = 0.5$	0.438	0.192	0.165	0.041	0.95	0.055	0.04	0.035	0.2	0	2.061	0.096	0.189	2.445	0
	$\alpha = -2$	-1.883	0.724	0.598	0.537	0.972	-	-	-	-	-	0.569	0.071	0.092	6.603	0
	$\gamma_0 = -0.5$	-0.498	0.188	0.188	0.035	0.953	-0.495	0.187	0.187	0.035	0.953	-	-	-	-	-
	$\gamma_1 = 1$	0.998	0.264	0.264	0.07	0.952	0.995	0.264	0.263	0.07	0.951	-	-	-	-	-
	mean_AIC			718.103					739.415					900.029		
	Percent_AIC			99%					0.5%					0. 5%		
N=500	$\beta_1 = 1$	0.998	0.144	0.139	0.021	0.946	1.084	0.134	0.095	0.025	0.903	1.681	0.217	0.155	0.51	0.123
	$\beta_2 = -0.5$	-0.506	0.113	0.116	0.013	0.961	-0.498	0.128	0.085	0.016	0.951	-0.483	0.201	0.138	0.041	0.953
	$\beta_1^* = 0.7$	0.623	0.229	0.234	0.058	0.951	0.398	0.146	0.132	0.112	0.448	0.921	0.153	0.144	0.073	0.717
	$\theta = 0.5$	0.453	0.139	0.126	0.022	0.948	0.064	0.028	0.03	0.191	0	2.076	0.067	0.119	2.489	0
	$\alpha = -2$	-1.728	0.402	0.399	0.236	0.903	-	-	-	-	-	0.562	0.05	0.056	6.567	0
	$\gamma_0 = -0.5$	-0.495	0.129	0.131	0.017	0.956	-0.494	0.13	0.116	0.017	0.953	-	-	-	-	-
	$\gamma_1 = 1$	0.991	0.183	0.185	0.034	0.951	0.986	0.183	0.163	0.034	0.951	-	-	-	-	-
	mean_AIC			1411.985					1465.70					1785.917		
	Percent_AIC			100%					0.00%					0.00%		
N=1000	$\beta_1 = 1$	0.999	0.102	0.099	0.01	0.952	2.21	1.14	2.289	1.31	0.98	1.697	0.157	0.072	0.51	0.014
	$\beta_2 = -0.5$	-0.505	0.083	0.082	0.007	0.96	-0.121	2.45	1.861	6.14	0.971	-0.46	0.145	0.064	0.022	0.957
	$\beta_1^* = 0.7$	0.616	0.159	0.164	0.032	0.913	1.77	9.51	1.972	91.6	0.971	0.916	0.106	0.067	0.058	0.475
	$\theta = 0.5$	0.471	0.098	0.094	0.011	0.945	0.064	0.023	0.002	0.19	0	2.083	0.048	0.055	2.507	0
	$\alpha = -2$	-1.665	0.261	0.277	0.18	0.736	-	-	-	-	-	0.565	0.035	0.026	6.581	0
	$\gamma_0 = -0.5$	-0.499	0.087	0.093	0.008	0.943	-0.023	1.68	0.284	2.89	0.968	-	-	-	-	-
	$\gamma_1 = 1$	0.993	0.129	0.131	0.017	0.945	1.22	1.61	0.469	2.65	0.971	-	-	-	-	-
	mean_AIC			2841.235					2948.954					3587.462		
	Percent_AIC			100%					0.00%					0.00%		

AIC, Akaike information criterion; CP, coverage probability; MSE, mean square error; SE, standard error

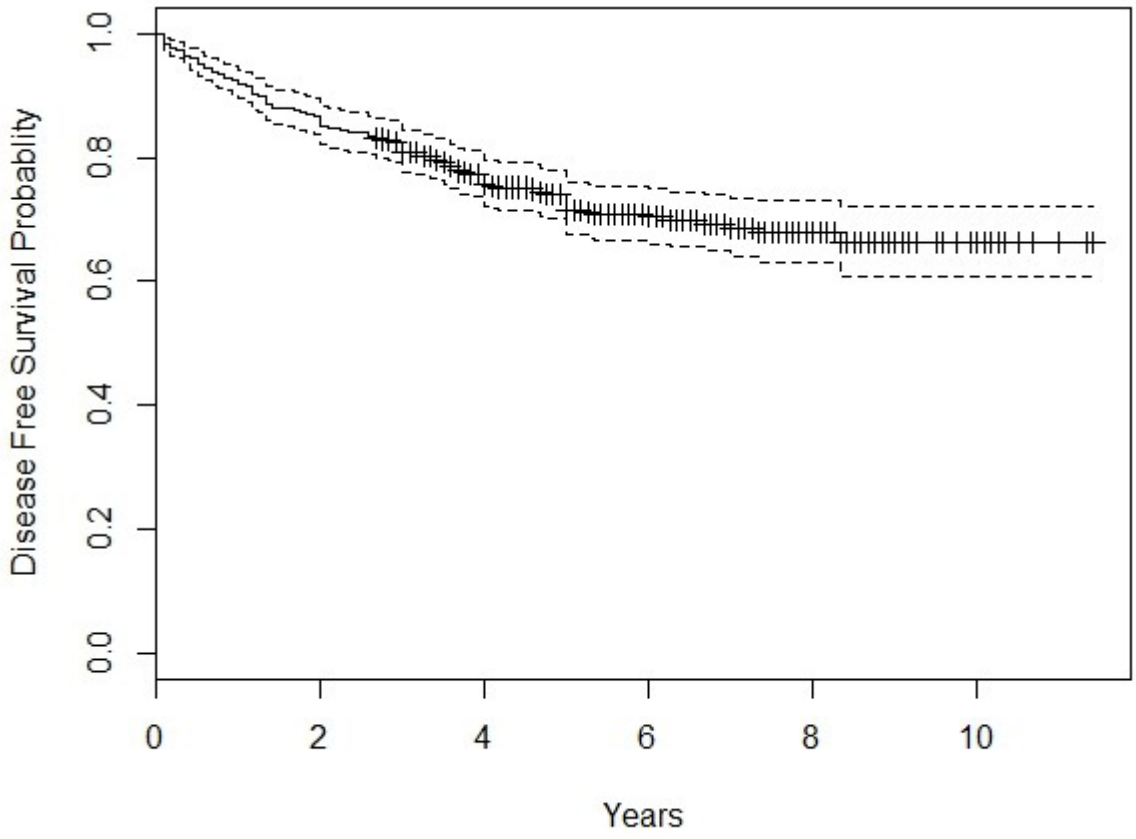
Table 5. *Application results*

Variables	Modalities	proposed model			Reduced Model 1 (Cure with same frailty)			Reduced Model 2 (Without cure with different frailty)		
		Est (SE)	HR	P-value	Est (SE)	HR	P-value	Est (SE)	HR	P-value
Recurrent events										
Age (ref:≤50)	>50	0.013(0.007)	1.013	0.64	0.014(0.008)	1.014	0.086	0.012(0.005)	1.012	0.018
Lymphovascular (ref:negative)	positive	0.008 (0.004)	1.008	0.067	0.008 (0.005)	1.008	0.091	0.007(0.003)	1.007	0.019
Lymph node Status(ref:negative)	positive	0.007 (0.004)	1.007	0.07	0.007 (0.004)	1.007	0.088	0.006(0.003)	1.006	0.023
Tumor size (ref:I)	II	0.012(0.007)	1.012	0.068	0.012(0.007)	1.012	0.076	0.01(0.004)	1.01	0.021
	III	0.008(0.005)	1.008	0.08	0.009(0.005)	1.009	0.083	0.006(0.002)	1.006	0.029
Cancer death										
Age (ref:≤50)	>50	0.375 (0.359)	1.45	0.231	0.355(0.346)	1.426	0.236	0.616(0.346)	1.852	0.081
Lymphovascular (ref:negative)	positive	0.938(0.503)	2.55	0.07	0.625(0.382)	1.868	0.105	0.419(0.374)	1.521	0.213
Lymph node status(ref:negative)	positive	0.621 (0.606)	1.89	0.236	-0.352 (0.477)	0.704	0.304	-0.624(0.449)	0.536	0.152
Tumor size (ref:I)	II	-0.725(0.446)	0.48	0.106	-0.259(0.442)	0.772	0.336	-1.024(0.438)	0.359	0.026
	III	-1.227(0.704)	0.29	0.87	-0.344(0.888)	0.709	0.37	-1.73(0.655)	0.177	0.012
Cure Logistic Model		Est (SE)	OR	P-value	Est (SE)	OR	P-value			
Intercept		0.867(0.742)	2.38	0.201	0.991(1.141)	2.693	0.274	-----	-----	-----
Tumor size	II	-0.651(0.657)	0.52	0.244	-1.067(0.833)	0.344	0.176	-----	-----	-----
	III	-1.47 (0.561)	0.23	0.013	-1.736 (1.059)	0.176	0.102	-----	-----	-----
θ =		0.904(0.332)	------	0.383	1.099(0.941)	------	0.397	1.288(0.173)	------	<0.1
α =		1.357(0.361)	------	<0.001	------	------	------	1.8(0.351)	------	<0.001
AIC		2062.394			2063.579			2080.42		

AIC, Akaike information criterion; HR, hazard ratio; SE, standard error

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Fig 1. Kaplan-Meier curve for the cancer free survival. The censoring time is denoted by "+".



ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Choosing Statistical Models to Assess Biological Interaction as a Departure from Additivity of Effects

David M. Thompson[(1)], *Yan Daniel Zhao*[(1)]

(1) Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104

CORRESPONDING AUTHOR: Yan Daniel Zhao, Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104. E-mail: daniel-zhao@ouhsc.edu, ORCID 0000-0003-3448-0527

SUMMARY

Vanderweele and Knol define biological interaction as an instance wherein "two exposures physically interact to bring about the outcome." A hallmark of biological interaction is that the total effect, produced when factors act together, differs from the sum of effects when the factors operate independently.
Epidemiologists construct statistical models to assess biological interaction. The form of the statistical model determines whether it is suited to detecting departures from additivity of effects or for detecting departures from multiplicativity of effects. A consensus exists that biological interaction should be assessed as a departure from additivity of effects.
This paper compares three statistical models' assessment of a data example that appears in several epidemiology textbooks to illustrate biological interaction in a binomial outcome. A linear binomial model quantifies departure from additivity in the data example in terms of differences in probabilities. It generates directly interpretable estimates and 95% confidence intervals for parameters including the interaction contrast (IC). Log binomial and logistic regression models detect no departure from multiplicativity in the data example. However, their estimates contribute to calculation of a "Relative Excess Risk Due to Interaction" (RERI), a measure of departure from additivity on a relative risk scale.
The linear binomial model directly produces interpretable assessments of departures from additivity of effects and deserves wider use in research and in the teaching of epidemiology. Strategies exist to address the model's limitations.

Keywords: additivity and multiplicativity of effects; biological interaction; statistical interaction; generalized linear models; interaction contrast (IC); Relative Excess Risk Due to Interaction (RERI)

INTRODUCTION

Biological interaction and statistical interaction

Hypotheses related to biological interaction are often of interest in studies of clinical or population health. Vanderweele and Knol [1] define biological interaction as an instance in which "two exposures physically interact to bring about the outcome." Rothman [2] states that "biologic interaction between two causes occurs whenever the effect of one is dependent on the presence of the other."

Investigators construct statistical models to detect interaction and effect modification. Rothman [2] points out that "in statistics, the term 'interaction' is used to refer to departure from the underlying form of a statistical model." A model's form can suit it for detecting departures from additivity of effects or for detecting departures from multiplicativity of effects. Because a statistical model's form affects the interpretation of statistical interaction, Rothman [2] prefers the term "effect measure modification" to interaction.

Rothman links "biological independence" with an additivity of effects and connects "biological interaction" with a departure from an additivity of effects. "Why is it," Rothman asks, "that biological interaction should be evaluated as departures from additivity of effect" [2]? By 2007, the STROBE statement regarded the

DOI: 10.54103/2282-0930/20180
Accepted: 4th August 2023

response to Rothman's rhetorical question to reflect a "consensus that the additive scale, which uses absolute risks, is more appropriate [than the multiplicative scale] for public health and clinical decision making" [3]. The authors of the STROBE statement remind investigators that "in many circumstances, the absolute risk associated with an exposure is of greater interest than the relative risk" and ask them to "consider translating estimates of relative risk into absolute risk for a meaningful time period" [3]. Vanderweele and Knol [1] remark, more pointedly, that "one reason why additive interaction is important to assess (rather than only relying on multiplicative interaction measures) is that it is the more relevant public health measure."

Additivity and multiplicativity of effects

This paper aligns with this consensus but avoids using the term "additive interaction." Instead, it links the concept to statistical models that assess evidence of a *departure from additivity of effects*. One such model, the "binomial model for the risk difference" [4], directly quantifies departures from additivity of effects in terms of differences in probabilities, including the interaction contrast (IC). This model is also called the "binomial regression model" [5, 6]. Richardson et al. [7], who employ it as a final step in a marginal structural model, call it the "linear binomial model," the term we will use.

In the linear binomial model, detection of statistical interaction constitutes direct evidence of a departure from additivity of effects. The log binomial and logistic regression models can also assess additivity indirectly, when their estimates of relative risks or odds ratios are recombined to calculate statistics like the "Relative Excess Risk due to Interaction" (RERI).

The paper also avoids using the term "multiplicative interaction" but links that concept to statistical models that assess evidence of *departures from multiplicativity of effects*. Log binomial models estimate effects in terms of relative risks, also called risk ratios, prevalence ratios [4,7] or prevalence proportion ratios. Logistic regression models estimate effects in terms of odds and odds ratios. In the log binomial and logistic models, which employ log transformations of probabilities or of their corresponding odds, detection of statistical interaction constitutes direct evidence of a departure from multiplicativity among effects.

METHODS

Statistical models for binomial outcomes

The linear binomial, log binomial and logistic regression models are all examples of generalized linear models. Each treats the outcome as arising from a binomial distribution. Each features a linear predictor structured as a *sum of terms*. In this regard, all generalized linear models might be considered "additive." Accordingly, this paper does not refer to "additive or multiplicative models" but refers instead to statistical models that assess additivity or multiplicativity of effects.

All three models link a binomial outcome to a linear predictor. They are distinguished by the link functions they employ. The linear binomial model uses the identity link, the log binomial model uses the log link, and the logistic regression model uses the logit link. Thus, the linear binomial model operates directly on probabilities, while the others apply log transformations of the probabilities or of their corresponding odds. Because each model estimates a different effect measure, they differ in their ability to detect statistical interaction in a collection of data.

After reviewing the definition of additivity of effects, we compare the three statistical models using a widely cited example of biological interaction [8]. The linear binomial model detects statistical interaction in these data. The log binomial and logistic regression models, which assess multiplicativity of relative risks or of odds ratios, find no evidence of statistical interaction. The absence of statistical interaction in these models does not point to an absence of biological interaction, but to a lack of departure from multiplicativity of effects.

We conclude by summarizing the three models' advantages and limitations for assessing additivity of effects. The RERI is commonly used in epidemiologic research to quantify departures from additivity despite complications in its estimation, testing and interpretation. In comparison, the linear binomial model produces readily interpretable estimates of effects, including the interaction contrast.

Defining additivity of effects

Consider a comparison of the probability or "risk" of an outcome Y among individuals who are exposed or not exposed to one or both of two "risk factors," X and Z. Then, pxz is a probability whose subscripts signify the probability or risk of the outcome Y at "levels" of X and Z (Table 1).

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

Table 1. Probabilities of an outcome (Y) at levels of two exposure or risk factors (X and Z)

	Z=1 ("exposed to factor Z")	Z=0 ("not exposed to factor Z")
X=1 ("exposed to factor X")	p_{11}	p_{10}
X=0 ("not exposed to factor X")	p_{01}	p_{00}

Rothman [2 (p.178)] states that the following equation "establishes additivity as the definition of biological independence."

$$p_{11} - p_{00} = (p_{10} - p_{00}) + (p_{01} - p_{00}) \qquad \text{(Equation 1)}$$

According to Rothman's equation, two exposures (X and Z) are biologically independent, and their effects are additive, when the effect on of their joint and simultaneous effects ($p_{11} - p_{00}$) is equal to the sum of the separate and independent effects of X ($p_{10} - p_{00}$) and of Z ($p_{01} - p_{00}$). A *departure* from additivity of effect, which Rothman considers evidence of biological interaction, is present when the exposures' joint and simultaneous effect differs from the sum of their separate effects.

Additivity can be defined equivalently as a homogeneity of effects. The terms of Equation 1 can be reordered to obtain

$$p_{11} - p_{01} = p_{10} - p_{00} \qquad \text{(Equation 2)}$$

$$p_{11} - p_{10} = p_{01} - p_{00} \qquad \text{(Equation 3)}$$

Equation 2 states that the effect of X on Y is the same whether Z = 1 ($p_{11} - p_{01}$) or Z = 0 ($p_{10} - p_{00}$). Homogeneity of effects is reciprocal. Equation 3 states that the effect of Z on Y is the same at all levels of X, that is, whether X=1 ($p_{11} - p_{10}$) or X=0 ($p_{01} - p_{00}$). When the effects of X and Z are additive, the association between Y and X is homogenous at levels of Z, and the association between Y and Z is homogenous at levels of X.

Assessing additivity of effects using probabilities (the interaction contrast) or ratios (the RERI)

Departures from an additivity of effects (or from biological independence), whether defined as an inequality between joint and independent effects, or as a heterogeneity among effects, can be formally assessed through the interaction contrast, whose terms are probabilities, and the RERI, whose terms are relative risks.

The terms in equation (1) can be ordered to produce the interaction contrast [9]:

$$p_{11} - p_{10} - p_{01} + p_{00} = 0 \qquad \text{(Equation 4)}$$

Reordering the terms in Equation 4 and dividing each by p_{00} yields:

$$p_{11}/p_{00} - p_{01}/p_{00} - p_{10}/p_{00} + 1 = 0.$$

Recognizing that these ratios of probabilities are relative risks (RR), we obtain:

$$RR_{11} - RR_{01} - RR_{10} + 1 = 0. \qquad \text{(Equation 5)}$$

Rothman [10] names the quantity on the left side of equation 5 the "Relative Excess Risk due to Interaction" (RERI). Rothman and Greenland [9] call it the "interaction contrast ratio" (ICR). Hosmer and Lemeshow [11] define it as "the proportion of disease among those with both exposures that is attributable to their interaction."

The algebraic equivalence between equations 1 (for the IC) and 5 (for the RERI) validates the assessment of additivity of effects on either probability or relative risk scales. The IC and the RERI formally test the hypothesis that the effects on Y of X and Z are additive or, equivalently, that no interaction exists between X and Z. The STROBE statement [3 (p.825)] illustrates how to use the RERI to assess departures from additivity of effects.

Data example. Lung cancer mortality among workers with different exposures to asbestos and smoking

Hammond et al. [8] compared the risk of a dichotomous outcome, mortality from lung cancer, among 17,800 asbestos workers and among 73,763 workers who were not exposed to asbestos. They also recorded smoking status, so participants displayed combinations of exposure to cigarette smoking and to asbestos (Table 2). Hammond's study is widely used in epidemiology textbooks [2 (pp.168-180),12] to illustrate biological interaction.

Supplementary File 1 illustrates the creation of a dataset that closely approximates the properties of the published data. So that the dataset's risk probabilities (reported as lung cancer deaths per 100,000) reflect the published ones, we assumed a smoking prevalence of 0.28 for both the asbestos workers and for the comparison group of unexposed workers.

Table 2. Lung cancer deaths (per 100,000 workers) among those with exposure to asbestos and/or cigarette smoking

	Asbestos Exposure	
Cigarette smoking	Asbestos Workers (n=17800)	Comparison Group (n=73763)
Smokers	p_{11}=601.9	p_{10}=121.1
Non-Smokers	p_{01}=54.6	p_{00}=11.3

The data example illustrates a departure from additivity of effects

If the effects of asbestos exposure and cigarette smoking are additive, the expected effect of experiencing both exposures would equal the sum of the exposures' separate effects (Equation 1). Following the notation introduced in Table 1 to define p_{xz}, where X denotes cigarette smoking (1 = smokers and 0 = nonsmokers) and Z denotes asbestos exposure (1 = exposed and 0 = not exposed), the estimated risk probabilities are:

$\hat{p}_{11} - \hat{p}_{00} = 601.9 - 11.3 = 590.6$ excess deaths per 100,000 people, attributable to joint effects of both exposures.

$\hat{p}_{10} - \hat{p}_{00} = 121.0 - 11.3 = 109.7$ excess deaths per 100,000 attributable to smoking by itself.

$\hat{p}_{01} - \hat{p}_{00} = 54.6 - 11.3 = 43.3$ excess deaths per 100,000 people, attributable to asbestos exposure by itself.

The number of lung cancer deaths attributable to dual exposure appears to exceed the sum of the exposures' separate effects. The interaction contrast for the data example: $p_{11} - p_{10} - p_{01} + p_{00}$ indicates that the risk of lung cancer death in those who experience both exposures exceeds, by about 437.6 deaths per 100,000, the sum of the separate risks from smoking or from asbestos exposure. Calculated for the data example, the RERI, which quantifies additivity of effects on the relative risk scale, $RR_{11} - RR_{01} - RR_{10} + 1$= [601.9/11.3] – [54.6/11.3] – [121.0/11.3] +1 = 38.7.

The linear binomial model directly estimates the interaction contrast in the data example

The linear binomial model [4,7] estimates the interaction contrast directly in terms of probabilities and differences in probabilities:

$$P(Y = 1) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ \qquad \text{(Equation 6)}$$

Recalling that X and Z take values of 1 for "exposure" and 0 for "no exposure", then

$$\hat{p}_{00} = \beta_0$$
$$\hat{p}_{10} - \hat{p}_{00} = (\beta_0 + \beta_1) - \beta_0 = \beta_1$$
$$\hat{p}_{01} - \hat{p}_{00} = (\beta_0 + \beta_2) - \beta_0 = \beta_2$$
$$\hat{p}_{11} - \hat{p}_{00} = (\beta_0 + \beta_1 + \beta_2 + \beta_3) - \beta_0 = \beta_1 + \beta_2 + \beta_3$$

Substituting these expressions into Equation 1, which defines additivity of effects,

$$p_{11} - p_{00} = (p_{10} - p_{00}) + (p_{01} - p_{00})$$
$$\beta_1 + \beta_2 + \beta_3 = \beta_1 + \beta_2$$

In the linear binomial model, effects are additive if β_3, the regression coefficient associated with the product or interaction term, is equal to zero.

Substituting the expressions into Equation 4 illustrates that the model's estimate for β_3 directly estimates the interaction contrast:

$$p_{11} - p_{10} - p_{01} + p_{00} = (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1) - (\beta_0 + \beta_2) + \beta_0 = \beta_3$$

Thus, the linear binomial model's estimates for the interaction contrast and for the X*Z interaction are equivalent. Both provide direct tests of additivity; evidence against the hypothesis that β_3 =0 is evidence of a departure from additivity.

Supplementary File 2 illustrates the construction of the linear binomial model using SAS PROC GENMOD [4,7]. The model's point estimates for the number of deaths per 100,000 workers, which are presented in Table 3, are equal to those reported in Table 2. Table 3 also reports the model's estimates (and 95% CI) for regression coefficients. These coefficients include estimates for the effect on lung cancer mortality of smoking among those not exposed to asbestos (β_1), and of asbestos exposure in non-smokers (β_2).

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

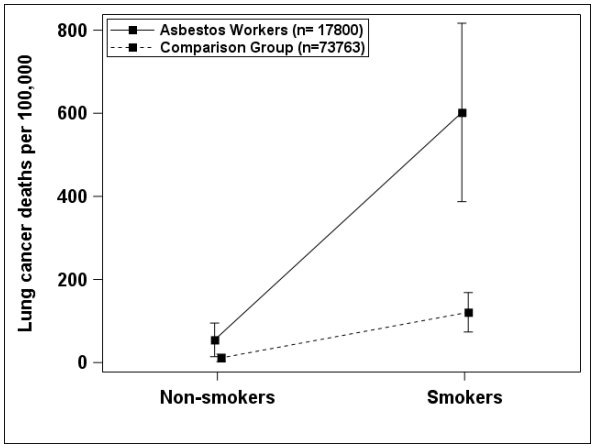
Table 3. Absolute risks (and risk differences) for death from lung cancer (per 100,000 workers) for those with exposure to asbestos and/or cigarette smoking, estimated by linear binomial model

	Smoking	Asbestos	Estimate	Deaths per 100,000	95% CI on estimate	
					Lower	Upper
p_{11}	1 (yes)	1 (yes)	0.006019	601.926	387.183	816.669
p_{10}	1 (yes)	0 (no)	0.001210	121.048	73.267	168.469
p_{01}	0 (no)	1 (yes)	0.000546	54.619	14.169	95.070
p_{00}	0 (no)	0 (no)	0.000113	11.298	2.258	20.337
β_1	smk ($\hat{p}_{10} - \hat{p}_{00}$)		0.001098	109.750	61.475	158.025
β_2	asbestos ($\hat{p}_{01} - \hat{p}_{00}$)		0.000433	43.322	1.873	84.770
β_3	smk*asbestos		0.004376	437.557	213.768	661.345
IC	$p_{11} - p_{10} - p_{01} + p_{00}$		0.004376	437.557	213.768	661.345

The linear binomial model produces identical inference for β_3, which estimates the statistical interaction between smoking and asbestos exposure, and for the IC (estimate: 437.6 deaths per 100,000; 95% CI: 213.8, 661.3; P=0.00012702). The consistency between the p values generated for these statistics verifies that they offer equivalent tests of the null hypothesis that the effects of smoking and asbestos exposure are additive.

Figure 1, which depicts the estimates and confidence intervals generated by the linear binomial model, illustrates the heterogeneity of the effects of smoking on lung cancer mortality in groups defined by asbestos exposure. The syntax that produced Table 3 and Figure 1 is contained in Supplementary File 3.

Figure 1. Biological interaction, between asbestos exposure and smoking, illustrated as a non-additivity or heterogeneity of effects



Log binomial and logistic regression models detect no departure from multiplicativity of effects in the data example

In contrast to the linear binomial model, models that employ logarithmic transformations of probabilities (log binomial models) or their corresponding odds (logistic regression models) assess departures from multiplicativity of effects. Multiplicativity of effects is defined in a manner analogous to the definition of additivity of effects. The effects of two factors (X and Z) on an outcome (Y) are multiplicative if their joint effects are equal to the *product* of their separate and independent effects. When effects are multiplicative, relative risks will conform to the relationship: $RR_{XZ} = RR_X \times RR_Z$, and odds ratios will conform to the relationship: $OR_{XZ} = OR_X \times OR_Z$. A log binomial model estimates and tests the multiplicativity of relative risks.

$$\ln[P(Y=1)] = \beta_0+\beta_1X+\beta_2Z+\beta_3XZ,$$

$$P(Y=1) = \exp(\beta_0+\beta_1X+\beta_2Z+\beta_3XZ).$$

it follows that

$$RR_{XZ} = \exp(\beta_0+\beta_1X+\beta_2Z+\beta_3XZ)$$
$$RR_X = \exp(\beta_1X)$$
$$RR_z = \exp(\beta_2Z)$$

If there is no departure from multiplicativity among relative risks, then:

$$RR_{XZ} = RR_XRR_Z$$

$$\exp(\beta_1X+\beta_2Z+\beta_3XZ) = \exp(\beta_1X)\exp(\beta_2Z) = \exp(\beta_1X+\beta_2Z)$$

These equalities hold only if β_3, the regression coefficient associated with the product term XZ, is equal to zero. Similarly, the logistic regression model, $\ln[P(Y=1) / P(Y=0)] = \beta_0+\beta_1X+\beta_2Z+\beta_3XZ$, assesses multiplicativity of effects expressed as odds or odds ratios. In either model, estimates or hypothesis tests that suggest that β_3 does not equal zero constitute evidence of a departure from multiplicativity of effects.

Applied to the data example, the log binomial model finds no evidence of statistical interaction between smoking and asbestos exposure (P=0.9637); measured as relative risks, the factors' effects are multiplicative and homogenous. Similarly, a logistic regression model finds no statistical interaction between smoking and asbestos exposure (P=0.9581) to suggest a departure from multiplicativity of effects measured as odds ratios. Figures 2 and 3 depict the estimates generated by the log binomial and logistic regression models. The models' construction, using SAS PROC GENMOD, is detailed in Supplementary File 4 along with the syntax that produced Figures 2 and 3.

Figure 2. Predicted log probabilities illustrate a lack of departure from multiplicativity of effects in the log binomial model.

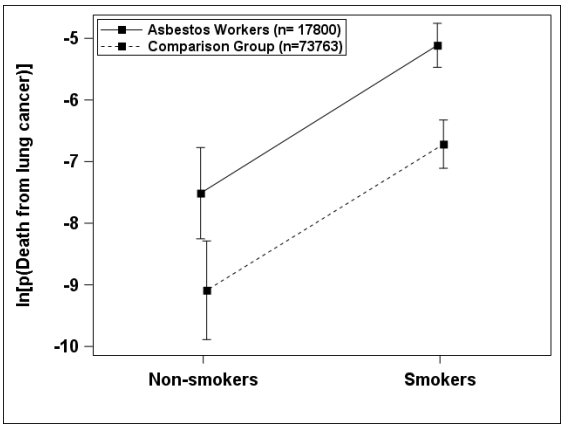
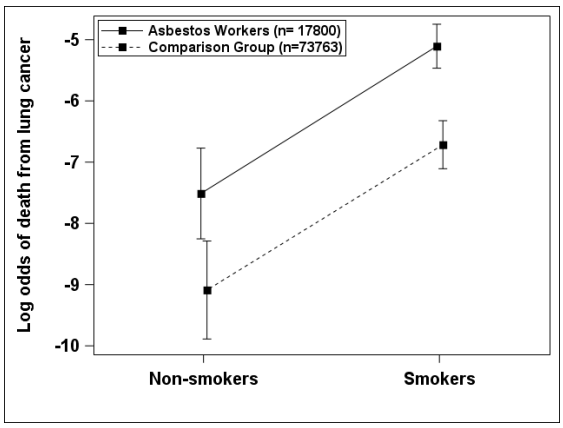


Figure 3. Predicted log odds illustrate a lack of departure from multiplicativity of effects in the logistic regression model.



The models' differences in detecting statistical interaction do not confound the question of whether the data exemplify biological interaction. Rather, they illustrate the importance of (1) identifying an effect measure (either a difference or a ratio between probabilities or risks) that reflects the hypothesized form of the interaction and then (2) constructing a statistical model that directly estimates that effect measure.

DISCUSSION

Choosing log binomial or logistic regression models that generate estimates of the RERI

Neither the log binomial model nor the logistic regression model detects statistical interaction in the data example. The models' form suits them for detecting departures from multiplicativity of effects. Nevertheless, they are widely used in epidemiology to assess departures from additivity of effects through ratio measures like the RERI [3].

Although widely used, the RERI has disadvantages. Because it is constructed from ratios, the RERI is not interpretable as the number of excess deaths attributable to exposure to both smoking and asbestos. The RERI of 38.7, calculated for the data example, lacks the ease of interpretation of the linear binomial model's estimate of the IC of 437.6 excess deaths per 100,000 (Table 3.) A second disadvantage relates to difficulties in obtaining standard errors with which to construct confidence intervals for or to test hypotheses related to the RERI. An influential approach, introduced by Hosmer and Lemeshow [11], estimates the RERI using logistic regression and obtains standard errors for its estimates using the delta method. SAS syntax for the approach is provided by Andersson et al. [13] and by Richardson and Kaufman [14], who construct a "linear odds ratio model" using SAS PROC NLMIXED. As an alternative approach, Richardson and Kaufmann [14] recommend bootstrapping for obtaining confidence intervals. An empirical 95% confidence interval on the RERI, calculated for the data example from 500 bootstrap samples, is 15.9, 132.6. However, because the bounds for the RERI's confidence interval are ratios, they present the same challenges to interpretation as the estimate itself.

Choosing the linear binomial model that directly estimates the interaction contrast

Logistic regression is widely used in epidemiology to study binomial outcomes, even though its form is suited for detecting departures from multiplicativity of effects. A major reason for the model's popularity and durability is that its use of the logit link, which is the canonical link for a binomial response, affords desirable statistical properties. Among these is logistic

regression's reliability in converging on parameter estimates. Models that use other link functions can encounter problems with convergence. Zou [15] and Spiegelman and Herzmark [4] discuss problems with convergence in the log binomial model and advocate use of a modified Poisson model to address the problem when it arises.

The linear binomial model, which uses the non-canonical identity link, can also fail to converge on estimates. This limitation interferes with the model's wider acceptance, despite its ability to directly assess additivity of effects by estimating the interaction contrast. To address non-convergence in the linear binomial model, Spiegelman and Herzmark [4] advocate modifying the model, retaining the identity link but assuming that the outcome follows a Poisson distribution. Although the approach ensures convergence, imposing the Poisson assumption causes the model to misspecify the binomial outcome's variance. This intentional misspecification of the outcome's distribution reduces the efficiency of the model's standard errors and of the hypothesis tests and confidence intervals that are based on them. Accordingly, Spiegelman and Herzmark [4] recommend calculating standard errors that are robust despite misspecification. Richardson et al. [7] also recommend the calculation of robust standard errors but, because they apply it to weighted data, do not advocate otherwise modifying the linear binomial model. Supplementary File 2 shows how to incorporate these various recommendations using SAS PROC GENMOD.

Cheung [5] addresses non-convergence in the linear binomial model by proposing a modified least squares (MLS) model that also uses the identity link. Cheung's approach also calculates robust standard errors. Cheung's approach differs in that it uses ordinary least squares (OLS) instead of maximum likelihood estimation (MLE). In doing so, it avoids specifying the outcome's assumed distribution. This strategy cures the problem of non-convergence but cannot guarantee that estimated probabilities will be in the logical range from 0 to 1.

CONCLUSION

Biological interaction is often hypothesized to manifest itself as a non-additivity of effects that are quantified as differences in risks or probabilities. Applied to a data example widely used in epidemiology education to illustrate biological interaction, a linear binomial model detects statistical interaction while logistic and log binomial models do not.

The result affirms the consensus that biological interaction should generally be assessed as a departure from an additivity of effects. Statistics like the RERI are widely used in epidemiology to assess additivity on a relative risk scale. In contrast, the linear binomial model produces estimates of differences in probabilities, including the interaction contrast, that are directly interpretable as excess risks.

Widely available software for generalized linear models permit researchers to construct the linear binomial model and to obtain estimates and confidence intervals for the interaction contrast and other effects. The model deserves wider use in research and judicious use in the teaching of epidemiology. The linear binomial model can encounter problems with convergence, but strategies exist to address this limitation.

FUNDING

Dr. Zhao's work was partially supported by funding provided by National Institutes of Health, National Institute of General Medical Sciences [Grant 1 U54GM104938, PI Judith James].

ACKNOWLEDGEMENTS

The authors thank Dr. Tabitha Garwe for important comments on the manuscript.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. VanderWeele TJ, Knol MJ. A tutorial on interaction. Epidemiologic Methods 2014;3:33-72.
2. Rothman KJ. Epidemiology: an introduction. New York: Oxford University Press, 2002.
3. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Epidemiology 2007;18(6):805-835.
4. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol 2005;162(3):199-200.
5. Cheung YB. A modified least-squares regression approach to the estimation of risk difference. Am J Epidemiol 2007;166(11):1337-44.
6. Bieler GS, Brown GG, Williams RL, Brogan DJ. Estimating model-adjusted risks, risk differences, and risk ratios from complex survey data. Am J Epidemiol 2010; 171(5):618-623.
7. Richardson DB, Kinlaw AC, MacLehose RF, Cole SR. Standardized binomial models for risk or prevalence ratios and differences. Int J Epidemiol 2015;44(5):1660-72.
8. Hammond EC, Selikoff IJ, Seidman H. Asbestos ex-

posure, cigarette smoking and death rates. Ann N Y Acad Sci 1979; 330:473-90.
9. Rothman KJ, Greenland S. Modern epidemiology. Philadelphia: Lippincott Williams and Wilkins, 1998.
10. Rothman KJ. Modern epidemiology (1st Ed.). Boston: Little, Brown and Company, 1986.
11. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. Epidemiology 1992; 3(5):452-456.
12. Szklo M, Nieto FJ. Epidemiology: beyond the basics (3rd Ed.). Sudbury, MA: Jones and Bartlett, 2004.
13. Andersson T, Alfredsson L, Källberg H, Zdravkovic S, Ahlbom, A. Calculating measures of biological interaction. Eur J Epidemiol 2005; 20(7):575-579.
14. Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds, Am J Epidemiol 2009; 16(6):756–760.
15. Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004;159(7):702-706.