BMJ Open

Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective big data observational cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030524
Article Type:	Protocol
Date Submitted by the Author:	19-Mar-2019
Complete List of Authors:	Zakowska, Izabela; Medical University of Lodz, Centre for Family and Community Medicine Kosiek, Katarzyna; Medical University of Lodz, Division of Public Health, Faculty of Medical Sciences Kowalczyk, Anna; Medical University of Lodz, Centre for Family and Community Medicine Grabowski, Jacek; Medical University of Lodz Godycki-Cwirko , Maciej; Medical University of Lodz, Centre for Family and Community Medicine; Medical University of Lodz, Division of Public Health, Faculty of Medical Sciences
Keywords:	elderly, COPD, local community, medical Big Data, COPD prevalence, exacerbations

Title page

Title of the article:

"Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective big data observational cohort study"

Izabela Zakowska[1], Katarzyna Kosiek[2], Anna Kowalczyk[1], Jacek Grabowski[3], Maciek Godycki-Cwirko[1, 2]

1. Centre for Family and Community Medicine, Medical University of Lodz, Kopcinskiego 20, 90-153 Lodz, Poland
2. Division of Public Health, Faculty of Medical Sciences, Medical University of Lodz, Plac Hallera 1, 90-647 Lodz, Poland
3. Medical University of Lodz, Lodz, Poland

Corresponding author:

Maciek Godycki-Cwirko
Centre for Family and Community Medicine, Medical University of Lodz, Kopcinskiego 20, 90-153 Lodz, Poland
maciej.godycki-cwirko@umed.lodz.pl
tel.: +48 42 679 55 46

Word count, excluding title page, abstract, references, figures and tables: 1603

Title: Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective big data observational cohort study

Abstract

Introduction

Analyses of large sets of electronic health related data (Big Data), including local community indicators, may improve knowledge on the outcomes of chronic diseases among patients and health care systems. Our study will estimate the prevalence of chronic obstructive pulmonary disease (COPD) and its exacerbations in elderly patients in the Lodz region, Poland, evaluate local community factors potentially associated with disease exacerbations and will rank the local communities according to health and local community indicators.

Methods and analysis

Local community factors, including medical/health, socioeconomic and environmental values potentially associated with COPD exacerbations, will be identified. A retrospective analysis will cover a cohort of about half a million people 65 years old and older, living in local communities of the Lodz region in 2016. Relevant data will be extracted from data bases, including those of the National Health Fund (NFZ), Tax Office (US) and National Statistics Center (GUS). The data will be checked for quality, cleaned and analyzed using data mining techniques. Logistic regression will be used to discover community determinants of COPD exacerbations in elderly patients.

Ethics and dissemination

The study protocol has been approved by the Bioethical Committee of Medical University of Lodz (RNN/248/18/KE, 10th July 2018). Our findings will be published in peer-reviewed

journals and reports. Recommendations will disseminated to key stakeholders including local leaders, decision makers, managers of prevention programs and local community media.

Strengths and limitations

- This will be a pioneering study of this kind in Poland to explore combined big sets of blinded health and local community data, extracted from the electronic databases of health-related records.
- The results will be visualized as maps.
- The main limitations relate to the specificity and sensitivity of the COPD coding, gaps in databases, short period of observation.

Key words: elderly, COPD, local community, medical Big Data, COPD prevalence, exacerbations

Abbreviations:

COPD - Chronic Obstructive Pulmonary Disease

ICD-10 - International Classification of Diseases

LC - Local Community (*gmina*)

LCS - Local Community Status

NFZ - National Health Fund

US - Tax Office (Revenue Administration Regional Office in Lodz)

GUS - National Statistics Center (Statistics Poland)

GP - general practice

EHRs - electronic health records

BS - brainstorming

FGD - Focus Group Discussion

Introduction

The incidence of chronic obstructive pulmonary disease, a non-reversible lung condition characterized by shortness of breath, chronic cough with sputum production, emphysema and systemic pulmonary inflammation, is increasing worldwide. [1] Its worldwide prevalence in adults has been estimated to be 10.1% [2]; it is currently the fourth leading cause of death worldwide and is predicted to become the third by 2020. [3]

The burden of the disease and its exacerbations has been studied globally from different perspectives. [4-11] Most studies to place a strong focus on clinical patterns, but others also consider the socioeconomic and environmental local community status of patients (LCS). [12 13] Community context has been identified as an important determinant of health outcomes. [14.] The term *community* here refers to a neighborhood and group of people living locally, within certain geographic construct boundaries, and is regarded as being synonymous with a *gmina* in Poland, defined by GUS as the basic unit of the three-tier territorial division of the country. The 177 *gminas*, and hence communities, included in our analysis constitute the Lodz voivodship, one of 16 such regions in the country.

Although certain predictors of exacerbations in COPD are well known [3 15-17], some community / regional factors are under examination. Pleasants *et al* conducted systematic review of the broad variety of factors to which patients are exposed in their living area. [13] The amount of data generated and collected routinely has increased significantly in the past decade, as has our ability to analyze and interpret it, especially in medicine. For example, a number of Big Data studies have been performed on Chinese health care, with large populations and the multiple structured and unstructured sources of data, with the aim of improving decision making. [18] It has been proposed that Big Data extracted by combining databases from various sources, including medical records, clinical and diagnostic results, patient medication records and medicine purchases, as well as data concerning costs, diagnostic costs and sports habits, could be used to improve the decision-making process, and thus influence patient health and quality of life. [19] Further analyses of large sets of electronic health records (including indicators of local communities, may improve knowledge about the outcomes of chronic diseases among patients.

Aims

1. To estimate the prevalence of COPD and its exacerbations in elderly patients living in the Lodz voivodship, Poland.
2. To evaluate local community factors potentially associated with disease exacerbations in this population.
3. To rank the *gminas* in the region according to health and local community indicators.

Our study is the pioneering of this kind in Poland, the purpose of which is to provide evidence for the potential role of local community factors in the health outcomes of the older population.

Methods and analysis

Study design

This will be a retrospective cohort study involving approximately half a million patients aged 65 years and older living in the Lodz voivodship, Poland, including patients with COPD and its exacerbations.

Data source

Data will be obtained from Big Dat**a** databases, such as National Health Fund with electronic health records of patients, the Tax Office and National Statistics Center).

Depersonalized data will be subjected to quality control and cleaning.

The scope of associations between the well-known patient-level risk factors and triggers of exacerbations of COPD, including local community factors will be identified with a literature review. Local community status factors will then be listed and selected with brainstorming (BS) and Focus Group Discussion (FGD), with the participation of researchers, experts and decision makers in the field of medicine and public health. During the BS and FGD, experts will select and classify factors into three main groups at *gmina* level, according to the 2015 Remington and Catlin methodology: 1) health factors 2) socioeconomic factors and 3) community environmental factors.[20] The group of experts will decide on the outline/framework of available databases and the collection of Big Data sets, and this outline will be filled with depersonalized data.

Population

Residents of the Lodz voivodship between 1st January and 31st December 2016, aged 65 and over will be identified from NFZ electronic health record systems, US and GUS using a residence code and assigned to a local community (*gmina*). Patients with COPD will be identified by the International Classification of Diseases (ICD-10) code J44 in their medical records; exacerbations will be defined as cases "hospitalized with the J44 code as a main reason for admission".

Study variables

This study will reveal a possible association between COPD exacerbations in elderly and local community factors: demographic, health care use, social, economic and environmental factors. It will take into account patient demographic and characteristics, including age, gender, residence code, as well as number of visits to the general practitioner (GP) in 2016, number of GP visits due to COPD in 2016, hospitalization, hospitalization with the J44 code as a main reason for admission, number of deaths, costs of care, patient income per *gmina* and number of GPs per *gmina*.

Patient and Public Involvement

Patients' priorities, experience, and preferences were obtained during previous European Union projects using the method such as Focus Group Discussion, were patient's suggestions to the COPD determinants were discussed and we took them into consideration.

Patients were not directly involved in the design of this study. As this is a protocol for a retrospective cohort study and no participant recruitment will take place, their involvement on the recruitment and dissemination of findings was not applicable. Result of the study will be available for public through internet and local media.

Analysis

The data obtained from the Big Data databases will be used to characterize patient health status, patient status related to the health care system, and the characteristics of the local community. Descriptive statistics for the total cohort, and the presence of COPD exacerbation will be calculated, aggregated at *gmina* level and categorized. Health characteristics and health outcomes will be aggregated by *gmina* in the Lodz voivodship, standardized and categorized.

Data mining techniques will be used to examine the relationships between patients and *gmina*; on the basis of which, indexes for each *gmina* will be calculated and normalized. Cross-sectional analysis and multivariable logistic regression (adjusted by demographics and health factors) will be used to test variables significantly associated with exacerbations of COPD. Health outcomes, such as the numbers of non-hospitalized patients with code J44 and numbers of those hospitalized (patients with exacerbation) within the *gmina*, will be categorized as a dependent variable in the regression analysis.

The obtained data will be visualized on a map of 177 *gminas* located in the Lodz region. Local community factors significantly associated with exacerbations of COPD will be shown on the *gmina* map according to each statistically significant factor [20]. The occurrence of exacerbations of COPD 65+ patients will be shown at *gmina* level using colors.

We will also calculate complex variables for each groups of determinants using the weights from the BS and FGD and literature review results. Additionally, the obtained complex variables related to health outcome and health determinants (health behaviors; clinical care; social and economic factors; and physical environment) for patients aged 65 and more with COPD exacerbations will also be visualized on the maps. It is planned therefore to obtain five maps, one for each of the five complex variables, illustrating the Lodz voivodship in terms of COPD exacerbations resolved at the *gmina* level.

All the data will be analyzed using the SAS 7.4 statistical package (SAS Institute, Cary, N.C., USA), MLwiN (Ver. 2.24; Centre for Multilevel Modelling, University of Bristol.), and STATISTICA 13.1.

Discussion

Our proposed methods will enable quantitative findings to be obtained that can be used to better understand the factors associated with exacerbations of COPD in communities.

The community-level contribution identified in the findings might be useful for future planning and resource allocation. This will be particularly useful if the obtained body of data is regularly updated by ongoing Big Data analysis of the *gminas* and health care systems.

Combining community and medical data can provide to key for informed recommendations for improving the quality of patient life in the local community.

This will be the pioneering study in Poland exploring combined sets of blinded health and local community Big Data, extracted from the electronic databases of health related records. The results will be visualized as maps.

Its main limitations relate to the specificity and sensitivity of the COPD coding, gaps in databases, short period of observation.

Acknowledgments

Special thanks to MSc. Edward Lowczowski for English language corrections.

Authors' contributions

The study concept and design was conceived by MGC, IZ, KK, AK and JG. Analysis will be performed by IZ and MGC. MGC, IZ, KK, AK, JG prepared the first draft of the manuscript. All authors provided edits and critiqued the manuscript for intellectual content.

Funding statement

This article was prepared within the research project no.2016/21/B/NZ7/02052 funded by Narodowe Centrum Nauki (National Science Centre Poland).

Competing interests statement

None declared.

Ethics and dissemination

The study protocol has been approved by the Bioethical Committee of Medical University of Lodz (RNN/248/18/KE, 10th July 2018). Our findings will be published in peer-reviewed journals and reports. Recommendations will disseminated to key stakeholders including local leaders, decision makers, managers of prevention programs and local community media.

References

1. Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. *Clin Chest Med* 2014;35(1):7-16. doi: 10.1016/j.ccm.2013.10.002 [published Online First: 2014/02/11]
2. Buist AS, McBurnie MA, Vollmer WM, et al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet* 2007;370(9589):741-50. doi: 10.1016/S0140-6736(07)61377-4 [published Online First: 2007/09/04]
3. Cardoso J, Coelho R, Rocha C, et al. Prediction of severe exacerbations and mortality in COPD: the role of exacerbation history and inspiratory capacity/total lung capacity ratio. *Int J Chron Obstruct Pulmon Dis* 2018;13:1105-13. doi: 10.2147/COPD.S155848 [published Online First: 2018/04/20]
4. GBD Chronic Respiratory Disease Collaborations, Soriano JB, Abajobir AA, et al. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017;5(9):691-706.
5. Donaldson GC, Seemungal TAR, Bhowmik A, et al. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax* 2002;57(10):847-52. doi: DOI 10.1136/thorax.57.10.847
6. Flattet Y, Garin N, Serratrice J, et al. Determining prognosis in acute exacerbation of COPD. *Int J Chron Obstruct Pulmon Dis* 2017;12:467-75.
7. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2095-128.
8. Seemungal TA, Donaldson GC, Paul EA, et al. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;157(5 Pt 1):1418-22.

9. Seemungal TAR, Hurst JR, Wedzicha JA. Exacerbation rate, health status and mortality in COPD--a review of potential interventions. *Int J Chron Obstruct Pulmon Dis* 2009;4:203-23.

10. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med* 2017;195(5):557-82.

11. Wedzicha JA, Seemungal TAR. COPD exacerbations: defining their cause and prevention. *Lancet* 2007;370(9589):786-96.

12. Grigsby M, Siddharthan T, Chowdhury MAH, et al. Socioeconomic status and COPD among low- and middle-income countries. *Int J Chronic Obstr* 2016;11:2497-507. doi: 10.2147/Copd.S111145

13. Pleasants RA, Riley IL, Mannino DM. Defining and targeting health disparities in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2016;11:2475-96.

14. Marmot MGB, M.; Davey Smith, G.;. Explanations for social inequalities in health. In: Amick IaB, and Levine, and S, and Tarlov, and A, R and Chapman, W and D, (eds.) ed. Society and health. New York: Oxford University Press 1995:172–210.

15. Chiba H, Abe S. [The environmental risk factors for COPD--tobacco smoke, air pollution, chemicals]. *Nihon Rinsho* 2003;61(12):2101-6.

16. Halpin DM, Miravitlles M, Metzdorf N, et al. Impact and prevention of severe exacerbations of COPD: a review of the evidence. *Int J Chron Obstruct Pulmon Dis* 2017;12:2891-908.

17. Viniol C, Vogelmeier CF. Exacerbations of COPD. *Eur Respir Rev* 2018;27(147)

18. Zhang L, Wang H, Li Q, et al. Big data and medical research in China. *Bmj* 2018;360:j5910.

19. Chen P-T. Medical big data applications: Intertwined effects and effective resource allocation strategies identified through IRA-NRM analysis. *Technological Forecasting & Social Change* 2018;130:150-64 doi: https://doi.org/10.1016/j.techfore.2018.01.033

20. Remington PL, Catlin BB, Gennuso KP. The County Health Rankings: rationale and methods. *Popul Health Metr* 2015;13:11.

BMJ Open

Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective Big Data observational cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030524.R1
Article Type:	Protocol
Date Submitted by the Author:	15-May-2019
Complete List of Authors:	Zakowska, Izabela; Medical University of Lodz, Centre for Family and Community Medicine Kosiek, Katarzyna; Medical University of Lodz, Division of Public Health, Faculty of Medical Sciences Kowalczyk, Anna; Medical University of Lodz, Centre for Family and Community Medicine Grabowski, Jacek; Medical University of Lodz Godycki-Cwirko, Maciek; Medical University of Lodz, Centre for Family and Community Medicine; Medical University of Lodz, Division of Public Health, Faculty of Medical Sciences
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Health services research
Keywords:	elderly, COPD, local community, medical Big Data, COPD prevalence, exacerbations

Title page

Title of the article:

"Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective Big Data observational cohort study"

Izabela Zakowska[1], Katarzyna Kosiek[2], Anna Kowalczyk[1], Jacek Grabowski[3], Maciek Godycki-Cwirko[1, 2]

1. Centre for Family and Community Medicine, Medical University of Lodz, Kopcinskiego 20, 90-153 Lodz, Poland
2. Division of Public Health, Faculty of Medical Sciences, Medical University of Lodz, Plac Hallera 1, 90-647 Lodz, Poland
3. Medical University of Lodz, Lodz, Poland

Corresponding author:

Maciek Godycki-Cwirko

Centre for Family and Community Medicine, Medical University of Lodz, Kopcinskiego 20, 90-153 Lodz, Poland

maciej.godycki-cwirko@umed.lodz.pl

tel.: +48 42 679 55 46

Word count, excluding title page, abstract, references, figures and tables: 2121

Title: Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective Big Data observational cohort study

Abstract

Introduction

Analyses of large sets of electronic health related data (Big Data), including local community indicators, may improve knowledge of the outcomes of chronic diseases among patients and healthcare systems. Our study will estimate the prevalence of chronic obstructive pulmonary disease (COPD) and its exacerbations in elderly patients in the Lodz region, Poland; it will also evaluate local community factors potentially associated with disease exacerbations and rank local communities according to health and local community indicators.

Methods and analysis

Local community factors, including medical/health, socioeconomic and environmental values potentially associated with COPD exacerbations will be identified. A retrospective analysis of a cohort of about half a million people 65 years old and older, living in local communities of the Lodz region in 2016 will be performed. Relevant data will be extracted from databases, including those of the National Health Fund (NFZ), Tax Office (US) and National Statistics Centre (GUS). This cross-sectional study will include data for a one-year period, from 1 January until 31 December 2016.

The data will first be checked for quality, cleaned and analyzed using data mining techniques, and then multilevel logistic regression will be used to discover the community determinants of COPD exacerbations.

Strengths and limitations

- This will be a pioneering study in Poland to explore combined big sets of blinded health and local community data extracted from the electronic databases of health-related records.
- The results will be visualized as maps.
- The main limitations relate to the specificity and sensitivity of the COPD coding, gaps in databases, short period of observation.
- The other limitation is the age limit of the patients.
- Additional limitations will be addressed when the study is completed.

Key words: elderly, COPD, local community, medical Big Data, COPD prevalence, exacerbations

Abbreviations:

COPD - Chronic Obstructive Pulmonary Disease

ICD-10 - International Classification of Diseases

LC - Local Community (*gmina*)

LCS - Local Community Status

NFZ - National Health Fund

US - Tax Office (Revenue Administration Regional Office in Lodz)

GUS - National Statistics Center (Statistics Poland)

GP - General Practice

EHRs - Electronic Health Records

BS - Brainstorming

FGD - Focus Group Discussion

Introduction

The world is currently seeing a growth in the incidence of chronic obstructive pulmonary disease, a non-reversible lung condition characterized by shortness of breath, chronic cough with sputum production, emphysema and systemic pulmonary inflammation. [1] Its worldwide prevalence in adults has been estimated to be 10% [2] and is among the leading causes of mortality and morbidity worldwide. [3]

The burden of the disease and its exacerbations has been studied globally from different perspectives. [4-10] The prevalence data are limited for Poland, where our study is located. Although the scope of the problem is well recognized worldwide, its impact seems to be poorly reflected in the Polish research data, so little is known of the exact prevalence of COPD in Poland.

Most studies place a strong focus on clinical patterns, but other also consider the socioeconomic and environmental local community status of patients (LCS). [11 12]

The community context has been identified as an important determinant of health outcomes. [13.] In the proposed study, the term *community* will be used to refer to a local neighborhood and its inhabitants within certain geographic construct boundaries, and is regarded as being synonymous with a *gmina* in Poland, defined by GUS as the basic unit of the three-tier territorial division of the country. The present analysis covers the whole of the Lodz voivodship, one of 16 in the country; it therefore comprises 177 *gminas*, i.e. communities.

Certain predictors of exacerbations in COPD are well known [14-17] while some community and regional factors remain under examination. Although a systematic review of the broad variety of factors to which patients are exposed in their living area has already been conducted by Pleasants *et al.* [12] the amount of data generated and collected routinely has increased significantly in the past decade, as has our ability to analyze and interpret it, especially in medicine. For example, a number of such Big Data studies have been performed using the Chinese healthcare system, including large populations and multiple structured and unstructured data sources, with the aim of improving decision making. [18] It has been proposed that Big Data extracted by combining databases from various sources, including medical records, clinical and diagnostic results, patient medication records and medicine purchases, as well as data concerning costs, diagnostic costs and sports habits, could be used to improve the decision-making process, and thus influence patient health and quality of life. [19] Further analyses of large sets of electronic health records, including indicators among local communities, may improve knowledge about the outcomes of chronic diseases among patients.

The members of the patient cohort will be COPD “labelled” patients who had been identified by the health care system and assigned the code J44 from International Classification of Diseases (ICD 10). We are aware of the limits of this approach, and that some COPD patients may not have been coded, and hence not included in the group, but our area of interest is the health care system dataset reflected by coding. A more detailed picture, and a more correct analysis, can be obtained by follow-up studies with more precise coding being applied and verified in the future.

Aims

1. To estimate the prevalence of J44 coded chronic obstructive pulmonary disease cases in elderly patients living in the Lodz voivodship, Poland.
2. To evaluate local community factors potentially associated with disease exacerbations in this population.
3. To rank the *gminas* in the region according to health and local community indicators.

Our study is the pioneering of this kind in Poland, the purpose of which is to provide evidence for the potential role of local community factors in the health outcomes of the older population.

Methods and analysis

Study design

This will be a retrospective cohort study involving approximately half a million patients aged 65 years and older living in the Lodz voivodship, Poland, including patients with COPD and its exacerbations. This study will include data for a one-year period, from 1 January until 31 December 2016.

Data source

Data will be obtained from Big Data databases, such as the electronic health records of patients from the NFZ (National Health Service), US (Tax Office) and GUS (National Statistics Center). Depersonalized data will be loaded and subjected to quality control and cleaning.

Individual patient data will be anonymized and assigned to the local communities which are the basic units of our analysis. We will collect three categories of data: (1) disease-related data, (2) health care services use-related data, (3) data relevant for selected local community indicators from restricted and publically-available databases and repositories with limited and unlimited access. Patient consent is not needed since we will not collect any personally-sensitive data.

Individual patient data will be matched by patient identifier within a single database. Individual data will not be matched between databases. Data will be matched on the local community level, and these matched local community data sets will be the units of our analyses.

The scope of associations between the well-known patient-level risk factors and triggers of exacerbations of COPD, including local community factors, will be identified with a literature review. Local community status factors will then be listed and selected with brainstorming (BS) and Focus Group Discussion (FGD), with the participation of researchers, experts and decision makers in the field of medicine and public health based on the methods described by Osborn [20 21] and Kitzinger [22], respectively. During the BS and FGD, experts will select and classify factors into three main groups at *gmina* level, according to the Remington and Catlin methodology, as follows: 1) health factors 2) socioeconomic factors and 3) community environmental factors. [23] The group of experts will decide on the outline/framework of available databases and the collection of Big Data sets, and this outline will be filled with depersonalized data.

Population

Residents of the Lodz voivodship aged 65 and over between 1st January and 31st December 2016 will be identified from NFZ electronic health record systems, US and GUS using a residence code and assigned to a local community (*gmina*). Patients with COPD will be identified by the International Classification of Diseases (ICD-10) code J44 in their medical records; exacerbations will be defined as cases "hospitalized with the J44 code as a main reason for admission".

Study variables

This study will reveal a possible association between COPD exacerbations in elderly and local community factors: demographic, health care use, social, economic and environmental factors. It will take into account patient demographic and characteristics, including age, gender, residence code, as well as the number of visits to the general practitioner (GP) in 2016, number of GP visits due to COPD in 2016, hospitalization, hospitalization with the J44 code as a main reason for admission, number of deaths, costs of care, patient income per *gmina* and number of GPs per *gmina*.

Patient and Public Involvement

The priorities, experience, and preferences of the patients and other health professionals were identified by individual interviews, BS and FGD technique in an earlier work. [24] The suggestions regarding COPD determinants were discussed and will be taken into consideration in the planned research.

Patients were not directly involved in the design of this study. As this is a protocol for a retrospective cohort study and no participant recruitment will take place, their involvement in the recruitment and dissemination of findings was not applicable.

The results of the study will be available for the public through internet and local media.

Statistical analysis

The data obtained from the Big Data databases will be used to characterize patient health status, patient status related to the health care system, and the characteristics of the local community. Descriptive statistics for the total group, and the presence of COPD exacerbation will be calculated, aggregated at *gmina* level and categorized. Health characteristics and health outcomes will be aggregated by *gmina* in the Lodz voivodship, standardized and categorized.

Data mining techniques will be used to examine the relationships between patients and *gmina*; on the basis of which, indexes for each *gmina* will be calculated and normalized. Cross-sectional, case-control multilevel multivariable logistic regression models (adjusted by demographics and health factors) will be used to test variables significantly associated with exacerbations of COPD. Health outcomes, such as the numbers of non-hospitalized patients awarded the code J44 and the numbers of hospitalized patients with exacerbation within the *gmina*, will be categorized as a dependent variable in the regression analysis.

The obtained data will be visualized on a map of 177 *gminas* located in the Lodz region. Local community factors significantly associated with exacerbations of COPD will be shown on the *gmina* map according to each statistically significant factor [23]. The occurrence of exacerbations of COPD 65+ patients will be shown at *gmina* level using colors.

Complex variables will be calculated for each group of determinants using the weights from the BS and FGD and literature review results. Additionally, the obtained complex variables related to health outcome and health determinants (health behaviors; clinical care; social and economic factors; and physical environment) for patients aged 65 and above with COPD exacerbations will also be visualized on the maps. It is planned therefore to obtain five maps, one for each of the five complex variables, illustrating the Lodz voivodship in terms of COPD exacerbations resolved at the *gmina* level.

All the data will be analyzed using the SAS 9.4 statistical package (SAS Institute, Cary, N.C., USA), MLwiN (Ver. 2.24; Centre for Multilevel Modelling, University of Bristol.) and STATISTICA 13.1.

Discussion

Our proposed methods will enable quantitative findings to be obtained that can be used to better understand the factors associated with exacerbations of COPD in communities.

The community-level contribution identified in the findings might be useful for future planning and resource allocation. This will be particularly useful if the obtained body of data is regularly updated by ongoing Big Data analysis of the *gminas* and healthcare systems.

Combining community and medical data can allow recommendations to be prepared for improving the quality of patient life in the local community.

This will be a pioneering study in Poland exploring combined sets of blinded health and local community Big Data, extracted from the electronic databases of health -related records. The results will be visualized as maps.

The main limitations relate to the specificity and sensitivity of the COPD coding, gaps in databases and short period of observation. We will select our study population based on the codes used by the national health service. We are aware of the bias related to this approach, such as limited code sets, mistaken coding, and errors related to the coding within the public datasets and repositories. Another limitation is fact that the 177 gminas were not randomly selected and were chosen based on their location within the Lodz voivodship.

Additional limitations and bias may be related to incompleteness and inaccuracy of data in databases; some variables of potential interest might not be available, as well as indicator selection might not be complete. The advantage is an ability to set a pilot framework for study disease in real world community environment.

Ethics and dissemination

The study protocol has been approved by the Bioethical Committee of Medical University of Lodz (RNN/248/18/KE, 10th July 2018). Our findings will be published in peer-reviewed journals and reports. Recommendations will disseminated to key stakeholders including local leaders, decision makers, managers of prevention programs and local community media.

Acknowledgments

Special thanks to mgr. Edward Lowczowski for English language corrections.

Authors' contributions

The study concept and design was conceived by MGC, IZ, KK, AK and JG. Analysis will be performed by IZ (SAS, MLwiN and STATISTICA statistical analyses) and MGC. MGC, IZ, KK, AK, JG prepared the first draft of the manuscript. All authors provided edits and critiqued the manuscript for intellectual content.

Funding statement

This article was prepared within the research project no.2016/21/B/NZ7/02052 funded by Narodowe Centrum Nauki (National Science Centre Poland).

Competing interests statement

None declared.

References

1. Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. *Clin Chest Med* 2014;35(1):7-16.
2. Buist AS, McBurnie MA, Vollmer WM, Gillespie S, Burney P, Mannino DM, et al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet* 2007;370(9589):741-50.
3. GBD C, Respiratory, Disease, Collaborators,. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017;5(9):691-706.

4. Donaldson GC, Seemungal TAR, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax* 2002;57(10):847-52.

5. Flattet Y, Garin N, Serratrice J, Perrier A, Stirnemann J, Carballo S. Determining prognosis in acute exacerbation of COPD. *Int J Chron Obstruct Pulmon Dis* 2017;12:467-75.

6. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2095-128.

7. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;157(5 Pt 1):1418-22.

8. Seemungal TA, Hurst JR, Wedzicha JA. Exacerbation rate, health status and mortality in COPD--a review of potential interventions. *Int J Chron Obstruct Pulmon Dis* 2009;4:203-23.

9. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med* 2017;195(5):557-82.

10. Wedzicha JA, Seemungal TAR. COPD exacerbations: defining their cause and prevention. *Lancet* 2007;370(9589):786-96.

11. Grigsby M, Siddharthan T, Chowdhury MAH, Siddiquee A, Rubinstein A, Sobrino E, et al. Socioeconomic status and COPD among low- and middle-income countries. *Int J Chronic Obstr* 2016;11:2497-507.

12. Pleasants RA, Riley IL, Mannino DM. Defining and targeting health disparities in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2016;11:2475-96.

13. Marmot MGB, M.; Davey Smith, G.;. Explanations for social inequalities in health. In: Amick IaB, and Levine, and S, and Tarlov, and A, R and Chapman, W and D, (eds.) editor. *Society and health.* New York: Oxford University Press, 1995:172–210.

14. Cardoso J, Coelho R, Rocha C, Coelho C, Semedo L, Bugalho Almeida A. Prediction of severe exacerbations and mortality in COPD: the role of exacerbation history and inspiratory capacity/total lung capacity ratio. *Int J Chron Obstruct Pulmon Dis* 2018;13:1105-13.

15. Chiba H, Abe S. [The environmental risk factors for COPD--tobacco smoke, air pollution, chemicals]. *Nihon Rinsho* 2003;61(12):2101-6.

16. Halpin DMG, Miravitlles M, Metzdorf N, Celli B. Impact and prevention of severe exacerbations of COPD: a review of the evidence. *Int J Chronic Obstr* 2017;12:2891-908.

17. Viniol C, Vogelmeier CF. Exacerbations of COPD. *Eur Respir Rev* 2018;27(147).

18. Zhang L, Wang H, Li Q, Zhao M-H, Zhan Q-M. Big data and medical research in China. *Bmj* 2018;360:j5910.

19. Chen P-T. Medical big data applications: Intertwined effects and effective resource allocation strategies identified through IRA-NRM analysis. *Technological Forecasting & Social Change* 2018;130:150-64

20. Osborn AF. *Applied imagination : principles and procedures of creative problem solving / by Alex F. Osborn.* 3d rev. ed. ed: New York : Charles Scribner's Sons, c1963., 1963.

21. Webpage. WHAT IS BRAINSTORMING AND HOW IS IT HELPFUL?, date accessed at 26.04.2019.

22. Kitzinger J. Qualitative research. Introducing focus groups. *BMJ* 1995;311(7000):299-302.

23. Remington PL, Catlin BB, Gennuso KP. The County Health Rankings: rationale and methods. *Popul Health Metr* 2015;13:11.

24. Krause J, Van Lieshout J, Klomp R, Huntink E, Aakhus E, Flottorp S, et al. Identifying determinants of care for tailoring implementation in chronic diseases: an evaluation of different methods. *Implement Sci* 2014;9:102.

BMJ Open

Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective Big Data observational cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030524.R2
Article Type:	Protocol
Date Submitted by the Author:	06-Jun-2019
Complete List of Authors:	Zakowska, Izabela; Medical University of Lodz, Centre for Family and Community Medicine Kosiek, Katarzyna; Medical University of Lodz, Division of Public Health, Faculty of Medical Sciences Kowalczyk, Anna; Medical University of Lodz, Centre for Family and Community Medicine Grabowski, Jacek; Medical University of Lodz Godycki-Cwirko, Maciek; Medical University of Lodz, Centre for Family and Community Medicine; Medical University of Lodz, Division of Public Health, Faculty of Medical Sciences
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Health services research
Keywords:	elderly, COPD, local community, medical Big Data, COPD prevalence, exacerbations

Title page

Title of the article:

"Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective Big Data observational cohort study"

Izabela Zakowska[1], Katarzyna Kosiek[2], Anna Kowalczyk[1], Jacek Grabowski[3], Maciek Godycki-Cwirko[1, 2]

1. Centre for Family and Community Medicine, Medical University of Lodz, Kopcinskiego 20, 90-153 Lodz, Poland
2. Division of Public Health, Faculty of Medical Sciences, Medical University of Lodz, Plac Hallera 1, 90-647 Lodz, Poland
3. Medical University of Lodz, Lodz, Poland

Corresponding author:

Maciek Godycki-Cwirko

Centre for Family and Community Medicine, Medical University of Lodz, Kopcinskiego 20,
90-153 Lodz, Poland

maciekgc@uni.lodz.pl

tel.: +48 42 679 55 46

Word count, excluding title page, abstract, references, figures and tables: 2141

Title: Community determinants of COPD exacerbations in elderly patients in Poland: protocol for a retrospective Big Data observational cohort study

Abstract

Introduction

Analyses of large sets of electronic health related data (Big Data), including local community indicators, may improve knowledge of the outcomes of chronic diseases among patients and healthcare systems. Our study will estimate the prevalence of chronic obstructive pulmonary disease (COPD) and its exacerbations in elderly patients in the Lodz region, Poland; it will also evaluate local community factors potentially associated with disease exacerbations and rank local communities according to health and local community indicators.

Methods and analysis

Local community factors, including medical/health, socioeconomic and environmental values potentially associated with COPD exacerbations will be identified. A retrospective analysis of a cohort of about half a million people 65 years old and older, living in local communities of the Lodz region in 2016 will be performed. Relevant data will be extracted from databases, including those of the National Health Fund, Tax Office and National Statistics Centre. This cross-sectional study will include data for a one-year period, from 1 January until 31 December 2016.

The data will first be checked for quality, cleaned and analyzed using data mining techniques, and then multilevel logistic regression will be used to discover the community determinants of COPD exacerbations.

Ethics and dissemination

The study protocol has been approved by the Bioethical Committee of Medical University of Lodz (RNN/248/18/KE, 10th July 2018). Our findings will be published in peer-reviewed journals and reports.

Strengths and limitations

- This will be a pioneering study in Poland to explore combined big sets of health and local community data extracted from the electronic databases.
- The results will be visualized as maps.
- The main limitations relate to the specificity and sensitivity of the COPD coding, gaps in databases, short period of observation.
- The other limitation is the age limit of the patients.
- Additional limitations will be addressed when the study is completed.

Key words: elderly, COPD, local community, medical Big Data, COPD prevalence, exacerbations

Abbreviations:

COPD - Chronic Obstructive Pulmonary Disease
ICD-10 - International Classification of Diseases
LC - Local Community (*gmina*)
LCS - Local Community Status
NFZ - National Health Fund
US - Tax Office (Revenue Administration Regional Office in Lodz)
GUS - National Statistics Center (Statistics Poland)
GP - General Practice
EHRs - Electronic Health Records
BS - Brainstorming

FGD - Focus Group Discussion

Introduction

The world is currently seeing a growth in the incidence of chronic obstructive pulmonary disease, a non-reversible lung condition characterized by shortness of breath, chronic cough with sputum production, emphysema and systemic pulmonary inflammation. [1] Its worldwide prevalence in adults has been estimated to be 10% [2] and is among the leading causes of mortality and morbidity worldwide. [3]

The burden of the disease and its exacerbations has been studied globally from different perspectives. [4-10] The prevalence data are limited for Poland, where our study is located. Although the scope of the problem is well recognized worldwide, its impact seems to be poorly reflected in the Polish research data, so little is known of the exact prevalence of COPD in Poland.

Most studies place a strong focus on clinical patterns, but other also consider the socioeconomic and environmental local community status of patients (LCS). [11 12]

The community context has been identified as an important determinant of health outcomes. [13.] In the proposed study, the term *community* will be used to refer to a local neighborhood and its inhabitants within certain geographic construct boundaries, and is regarded as being synonymous with a *gmina* in Poland, defined by GUS as the basic unit of the three-tier territorial division of the country. The present analysis covers the whole of the Lodz voivodship, one of 16 in the country; it therefore comprises 177 *gminas*, i.e. communities.

Certain predictors of exacerbations in COPD are well known [14-17] while some community and regional factors remain under examination. Although a systematic review of the broad variety of factors to which patients are exposed in their living area has already been conducted by Pleasants *et al.* [12] the amount of data generated and collected routinely has increased significantly in the past decade, as has our ability to analyze and interpret it, especially in medicine. For example, a number of such Big Data studies have been performed using the Chinese healthcare system, including large populations and multiple structured and unstructured data sources, with the aim of improving decision making. [18] It has been proposed that Big Data extracted by combining databases from various sources, including medical records, clinical and diagnostic results, patient medication records and medicine purchases, as well as data concerning costs, diagnostic costs and sports habits, could be used to improve the decision-making process, and thus influence patient health and quality of life. [19] Further analyses of large sets of electronic health records, including indicators among local communities, may improve knowledge about the outcomes of chronic diseases among patients.

The members of the patient cohort will be COPD “labelled” patients who had been identified by the health care system and assigned the code J44 from International Classification of Diseases (ICD 10). We are aware of the limits of this approach, and that some COPD patients may not have been coded, and hence not included in the group, but our area of interest is the health care system dataset reflected by coding. A more detailed picture, and a more correct analysis, can be obtained by follow-up studies with more precise coding being applied and verified in the future.

Aims

1. To estimate the prevalence of J44 coded chronic obstructive pulmonary disease cases in elderly patients living in the Lodz voivodship, Poland.
2. To evaluate local community factors potentially associated with disease exacerbations in this population.
3. To rank the *gminas* in the region according to health and local community indicators.

Our study is the pioneering of this kind in Poland, the purpose of which is to provide evidence for the potential role of local community factors in the health outcomes of the older population.

Methods and analysis

Study design

This will be a retrospective cohort study involving approximately half a million patients aged 65 years and older living in the Lodz voivodship, Poland, including patients with COPD and its exacerbations. This study will include data for a one-year period, from 1 January until 31 December 2016. The study reported in the manuscript (data extraction and analysis) will take place from 10 July 2018 until the 29 February 2020.

Data source

Data will be obtained from Big Data databases, such as the electronic health records of patients from the NFZ (National Health Service), US (Tax Office) and GUS (National Statistics Center). Depersonalized data will be loaded and subjected to quality control and cleaning.

Individual patient data will be anonymized and assigned to the local communities which are the basic units of our analysis. We will collect three categories of data: (1) disease-related data, (2) health care services use-related data, (3) data relevant for selected local community indicators from restricted and publically-available databases and repositories with limited and unlimited access. Patient consent is not needed since we will not collect any personally-sensitive data.

Individual patient data will be matched by patient identifier within a single database. Individual data will not be matched between databases. Data will be matched on the local community level, and these matched local community data sets will be the units of our analyses.

The scope of associations between the well-known patient-level risk factors and triggers of exacerbations of COPD, including local community factors, will be identified with a literature review. Local community status factors will then be listed and selected with brainstorming (BS) and Focus Group Discussion (FGD), with the participation of researchers, experts and decision makers in the field of medicine and public health based on the methods described by Osborn [20 21] and Kitzinger [22], respectively. During the BS and FGD, experts will select and classify factors into three main groups at *gmina* level, according to the Remington and Catlin methodology, as follows: 1) health factors 2) socioeconomic factors and 3) community environmental factors. [23] The group of experts will decide on the outline/framework of available databases and the collection of Big Data sets, and this outline will be filled with depersonalized data.

Population

Residents of the Lodz voivodship aged 65 and over between 1st January and 31st December 2016 will be identified from NFZ electronic health record systems, US and GUS using a residence code and assigned to a local community (*gmina*). Patients with COPD will be identified by the International Classification of Diseases (ICD-10) code J44 in their medical records; exacerbations will be defined as cases "hospitalized with the J44 code as a main reason for admission".

Study variables

This study will reveal a possible association between COPD exacerbations in elderly and local community factors: demographic, health care use, social, economic and environmental factors. It will take into account patient demographic and characteristics, including age, gender, residence code, as well as the number of visits to the general practitioner (GP) in 2016, number of GP visits due to COPD in 2016, hospitalization, hospitalization with the J44 code as a main reason for admission, number of deaths, costs of care, patient income per *gmina* and number of GPs per *gmina*.

Patient and Public Involvement

The priorities, experience, and preferences of the patients and other health professionals were identified by individual interviews, BS and FGD technique in an earlier work. [24] The suggestions regarding COPD determinants were discussed and will be taken into consideration in the planned research.

Patients were not directly involved in the design of this study. As this is a protocol for a retrospective cohort study and no participant recruitment will take place, their involvement in the recruitment and dissemination of findings was not applicable.

The results of the study will be available for the public through internet and local media.

Statistical analysis

The data obtained from the Big Data databases will be used to characterize patient health status, patient status related to the health care system, and the characteristics of the local community. Descriptive statistics for the total group, and the presence of COPD exacerbation will be calculated, aggregated at *gmina* level and categorized. Health characteristics and health outcomes will be aggregated by *gmina* in the Lodz voivodship, standardized and categorized.

Data mining techniques will be used to examine the relationships between patients and *gmina*; on the basis of which, indexes for each *gmina* will be calculated and normalized. Cross-sectional, case-control multilevel multivariable logistic regression models (adjusted by demographics and health factors) will be used to test variables significantly associated with exacerbations of COPD. Health outcomes, such as the numbers of non-hospitalized patients awarded the code J44 and the numbers of hospitalized patients with exacerbation within the *gmina*, will be categorized as a dependent variable in the regression analysis.

The obtained data will be visualized on a map of 177 *gminas* located in the Lodz region. Local community factors significantly associated with exacerbations of COPD will be shown on the *gmina* map according to each statistically significant factor [23]. The occurrence of exacerbations of COPD 65+ patients will be shown at *gmina* level using colors.

Complex variables will be calculated for each group of determinants using the weights from the BS and FGD and literature review results. Additionally, the obtained complex variables related to health outcome and health determinants (health behaviors; clinical care; social and economic factors; and physical environment) for patients aged 65 and above with COPD exacerbations will also be visualized on the maps. It is planned therefore to obtain five maps, one for each of the five complex variables, illustrating the Lodz voivodship in terms of COPD exacerbations resolved at the *gmina* level.

All the data will be analyzed using the SAS 9.4 statistical package (SAS Institute, Cary, N.C., USA), MLwiN (Ver. 2.24; Centre for Multilevel Modelling, University of Bristol.) and STATISTICA 13.1.

Discussion

Our proposed methods will enable quantitative findings to be obtained that can be used to better understand the factors associated with exacerbations of COPD in communities.

The community-level contribution identified in the findings might be useful for future planning and resource allocation. This will be particularly useful if the obtained body of data is regularly updated by ongoing Big Data analysis of the *gminas* and healthcare systems.

Combining community and medical data can allow recommendations to be prepared for improving the quality of patient life in the local community.

This will be a pioneering study in Poland exploring combined sets of blinded health and local community Big Data, extracted from the electronic databases of health -related records. The results will be visualized as maps.

The main limitations relate to the specificity and sensitivity of the COPD coding, gaps in databases and short period of observation. We will select our study population based on the codes used by the national health service. We are aware of the bias related to this approach, such as limited code sets, mistaken coding, and errors related to the coding within the public datasets and repositories. Another limitation is fact that the 177 gminas were not randomly selected and were chosen based on their location within the Lodz voivodship.

Additional limitations and bias may be related to incompleteness and inaccuracy of data in databases; some variables of potential interest might not be available, as well as indicator selection might not be complete. The advantage is an ability to set a pilot framework for study disease in real world community environment.

Ethics and dissemination

The study protocol has been approved by the Bioethical Committee of Medical University of Lodz (RNN/248/18/KE, 10th July 2018). Our findings will be published in peer-reviewed journals and reports. Recommendations will disseminated to key stakeholders including local leaders, decision makers, managers of prevention programs and local community media.

Acknowledgments

Special thanks to mgr. Edward Lowczowski for English language corrections.

Authors' contributions

The study concept and design was conceived by MGC, IZ, KK, AK and JG. Analysis will be performed by IZ (SAS, MLwiN and STATISTICA statistical analyses) and MGC. MGC, IZ, KK, AK, JG prepared the first draft of the manuscript. All authors provided edits and critiqued the manuscript for intellectual content.

Funding statement

This article was prepared within the research project no.2016/21/B/NZ7/02052 funded by Narodowe Centrum Nauki (National Science Centre Poland).

Competing interests statement

None declared.

References

1. Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. *Clin Chest Med* 2014;35(1):7-16.

2. Buist AS, McBurnie MA, Vollmer WM, Gillespie S, Burney P, Mannino DM, et al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet* 2007;370(9589):741-50.

3. GBD C, Respiratory, Disease, Collaborators,. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic

obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017;5(9):691-706.

4. Donaldson GC, Seemungal TAR, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax* 2002;57(10):847-52.

5. Flattet Y, Garin N, Serratrice J, Perrier A, Stirnemann J, Carballo S. Determining prognosis in acute exacerbation of COPD. *Int J Chron Obstruct Pulmon Dis* 2017;12:467-75.

6. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2095-128.

7. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;157(5 Pt 1):1418-22.

8. Seemungal TA, Hurst JR, Wedzicha JA. Exacerbation rate, health status and mortality in COPD--a review of potential interventions. *Int J Chron Obstruct Pulmon Dis* 2009;4:203-23.

9. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med* 2017;195(5):557-82.

10. Wedzicha JA, Seemungal TAR. COPD exacerbations: defining their cause and prevention. *Lancet* 2007;370(9589):786-96.

11. Grigsby M, Siddharthan T, Chowdhury MAH, Siddiquee A, Rubinstein A, Sobrino E, et al. Socioeconomic status and COPD among low- and middle-income countries. *Int J Chronic Obstr* 2016;11:2497-507.

12. Pleasants RA, Riley IL, Mannino DM. Defining and targeting health disparities in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2016;11:2475-96.

13. Marmot MGB, M.; Davey Smith, G.;. Explanations for social inequalities in health. In: Amick IaB, and Levine, and S, and Tarlov, and A, R and Chapman, W and D, (eds.) editor. *Society and health.* New York: Oxford University Press, 1995:172–210.

14. Cardoso J, Coelho R, Rocha C, Coelho C, Semedo L, Bugalho Almeida A. Prediction of severe exacerbations and mortality in COPD: the role of exacerbation history and inspiratory capacity/total lung capacity ratio. *Int J Chron Obstruct Pulmon Dis* 2018;13:1105-13.

15. Chiba H, Abe S. [The environmental risk factors for COPD--tobacco smoke, air pollution, chemicals]. *Nihon Rinsho* 2003;61(12):2101-6.

16. Halpin DMG, Miravitlles M, Metzdorf N, Celli B. Impact and prevention of severe exacerbations of COPD: a review of the evidence. *Int J Chronic Obstr* 2017;12:2891-908.

17. Viniol C, Vogelmeier CF. Exacerbations of COPD. *Eur Respir Rev* 2018;27(147).

18. Zhang L, Wang H, Li Q, Zhao M-H, Zhan Q-M. Big data and medical research in China. *Bmj* 2018;360:j5910.

19. Chen P-T. Medical big data applications: Intertwined effects and effective resource allocation strategies identified through IRA-NRM analysis. *Technological Forecasting & Social Change* 2018;130:150-64

20. Osborn AF. *Applied imagination : principles and procedures of creative problem solving / by Alex F. Osborn.* 3d rev. ed. ed: New York : Charles Scribner's Sons, c1963., 1963.

21. Web-page. WHAT IS BRAINSTORMING AND HOW IS IT HELPFUL?

22. Kitzinger J. Qualitative research. Introducing focus groups. *BMJ* 1995;311(7000):299-302.

23. Remington PL, Catlin BB, Gennuso KP. The County Health Rankings: rationale and methods. *Popul Health Metr* 2015;13:11.

24. Krause J, Van Lieshout J, Klomp R, Huntink E, Aakhus E, Flottorp S, et al. Identifying determinants of care for tailoring implementation in chronic diseases: an evaluation of different methods. *Implement Sci* 2014;9:102.