Dialogues in Health 6 (2025) 100216

Contents lists available at ScienceDirect

Dialogues in Health

journal homepage: www.elsevier.com/locate/dialog

Formulae for calculating subgroup disease burden from overall and reported or assumed relative burden estimates

Kwame Owusu-Edusei [a,*], Arijita Deb [b], Elamin H. Elbasha [a]

[a] Merck & Co., Inc., Rahway, NJ, USA
[b] GSK, Upper Province, PA, USA

A B S T R A C T

Background: The risk of disease varies across populations based on factors like age, sex, race, ethnicity, socioeconomic status, and underlying medical conditions. Subgroup or subpopulation data are critical in planning, executing and evaluating public health interventions. However, most studies report the values for the overall (total) population with little or no information on the subgroups. As a result, finding subgroup specific data can be challenging.
Objective: In this report, a set of formulae that can be used to calculate subgroup or subpopulation data using the overall estimates and the reported or assumed relative estimates were derived.
Methods: A simple numerical example was used to illustrate the methodology. Next, symbolic formula for calculating the burden (e.g., incidence, prevalence, or average cost) for 3 (and extended to *n* number of) subgroups or subpopulations were derived. To account for uncertainty in the data, two statistical methods were used to estimate confidence intervals for the point estimates.
Results: The derived formulae indicated that each subgroup or subpopulation's burden (incidence, prevalence, or average cost) can be calculated as the overall burden adjusted by the ratio of that subgroup or subpopulation's relative burden to the sum of the proportion-weighted relative burden (incidence, prevalence, or average cost) of all the subgroups or subpopulations within the population.
Conclusion: These formulae can help to avoid or minimize potential quantitative and qualitative errors in subgroup or subpopulation disease burden estimates used for health research, interventions and/or policy analyses or deliberations.

1. Introduction

To successfully initiate, plan, execute, assess, and modify public health interventions, timely collection of reliable data on disease burden (which includes both the health, as measured in terms of incidence and prevalence, and economic consequences such as average cost) over time for the population of interest is critical. The risk and burden of disease varies across populations based on factors like age, sex, race, ethnicity, socioeconomic status, and underlying medical conditions. For example, disease rates in many countries vary significantly by age, with infants and the elderly facing the highest risks. In a systematic review of 32 studies, Cong et al., [1] reported that hospitalization associated with respiratory syncytial virus (RSV) showed a U-shape age pattern with the lowest RSV hospitalization rates found in 5–17 years and highest rates found in <1 year and the elderly. Because the overall disease metric for any population is a direct reflection of the burden among subgroups or subpopulations, it is equally important to track and assess the health and economic burden (i.e., incidence, prevalence, or average cost) for the subgroups or subpopulations within the population. For some diseases (such as infectious diseases), the effectiveness of the public health intervention or strategy are largely dependent on its impact on specific subgroups or subpopulations within the population. In addition, accurate tracking of subgroup or subpopulation disease burden is pivotal in setting goals as well as achieving and/or improving health equity—Healthy People 2030 [2].

For one or a combination of reasons (such as study focus, research question or lack of resources) most data generated on the burden of disease focus on the entire population. As a result, data on specific subgroups (for use as inputs in cost-effectiveness analysis [CEAs] or budget impact analyses [BIAs]) may not be available. To further illustrate these data challenges, an example in pneumococcal disease (PD) research and interventions is presented in the subsequent section using health risk subgroups.

1.1. Risk-specific burden estimates for PD

Being the leading cause of pneumonia worldwide, *Streptococcus pneumoniae* causes invasive and non-invasive PD [3]. Invasive pneumococcal disease (IPD) includes osteomyelitis, bacteremia without focus, bacteremic pneumonia, septic arthritis and meningitis [3]. Non-

* Corresponding author at: Biostatistics & Research Decision Sciences, Merck & Co., Inc., WP37A-150, PO Box 1000, West Point, PA 19486, Rahway, NJ, USA.
E-mail address: kwame.owusu-edusei@merck.com (K. Owusu-Edusei).

https://doi.org/10.1016/j.dialog.2025.100216
Received 2 January 2025; Received in revised form 29 March 2025; Accepted 15 April 2025
Available online 18 April 2025

invasive disease include non-bacteremic pneumococcal pneumonia (NBPP), otitis media, and sinusitis [3]. NBPP (inpatient and outpatient manifestations) accounts for approximately 75 % of cases of pneumococcal pneumonia [4].

The introduction of new vaccines designed to protect against PD (Vaxneuvance ™ [Merck Sharp & Dohme LLC, a subsidiary of Merck & Co., Inc., Rahway, NJ, USA] and Prevnar 20 ™ [Wyeth Pharmaceuticals LLC, a subsidiary of Pfizer Inc., New York, NY, USA]) required population-level health and economic assessment of the new vaccines on the adult population using CEAs, which required data on the burden of PD—IPD, inpatient and outpatient NBPP [5]. However, while there are surveillance systems in several countries that track nationally representative IPD burden (such as the Active Bacteria Core Surveillance [ABCs] [6,7] in the United States [US] and the Surveillance Atlas of Infectious Diseases [8] in the European Union [EU]), similar nationally representative data on NBPP burden is lacking, due largely to the challenges and limitations associated with diagnostic tests [9].

In the US, existing data on NBPP burden have been estimated primarily from administrative health insurance claims databases [10,11]. Because there are substantial differences in the burden as well as the prophylactic or therapeutic responses (and associated costs) that depend on patient risk profiles, CEAs have included risk-specific inputs to account for the differences in the population-level health and economic outcomes. As a result, risk-specific burden estimates are important for health and economic evaluations of preventive and control measures [12–25]. While the NBPP burden for the entire population may be available through database analyses, subgroup- or subpopulation-level data are lacking.

1.2. Data/methods gap

Beside the data gap for subgroups or subpopulations, a comprehensive search in the literature for ways to estimate/calculate subgroup burden when the actual data is not available (or cannot be analyzed due to access restrictions and/or time constraints) yielded no results. Results from the literature search showed that studies on "subgroup analysis" or similar terms were focused on analyzing available data to obtain subgroup or subpopulation metrics. Consequently, this data/methods gap prompted the following question—given overall population metrics (such as incidence), subgroup profiles (such as risk status and distribution) and plausible assumptions about the relative risks, can the metrics (such incidence) for subgroups or subpopulations be calculated?

1.3. Objectives

Given the data question in the preceding paragraph, the objectives of this study are twofold:

1. Use an example to set up associated equations and calculate the corresponding incidence for each subgroup.
2. Build on the solution to objective 1 and derive symbolic formulae for the three major (and commonly used) population-level measures of disease burden—incidence (i.e., number of new cases over a specified time [26]), prevalence (i.e., number of new and existing cases over a specified time [26]), and average treatment cost, then extend to *n* subgroups.

The formulae derived in this report, which is based on plausible assumptions about relative burden or value, can help to obtain more realistic quantitative and qualitative estimates for epidemiologic and economic assessments of the various subgroups as well as the entire population of interest.

2. Methods and results

2.1. Data question

Details of the data question for outpatient NBPP incidence are presented in Table 1. For persons aged 18—49 years, the overall outpatient NBPP incidence in 2014 was reported by Tong et al., 2018. [11]. However, the outpatient NBPP incidence data was not available by risk group. As a result, it was assumed that the relative risk or incidence rate ratio and the risk distribution (percent in each risk group) were the same as those reported by Pelton et al., 2019 for all-cause pneumonia during the period 2013–2015 [10].

The assumption of same relative risk or incidence rate ratio and risk distribution (percent in each risk group) was made for three reasons: First, the two studies examined the same diseases (pneumonia-related outcomes) using similar ICD-9 (International Classification of Diseases 9th edition) and HCPCS (Healthcare Common Procedure Coding System) codes. Second, the two observational studies used the same database (Truven Health MarketScan® Commercial Claims and Encounters). Third, the two studies analyzed data from similar overlapping periods 2008–2014 [11] and 2007–2015 [10] and on similar age groups.

Based on their (Pelton et al., 2018 [10]) risk profile results, adults who were immunocompromised or had a cochlear implant were categorized as High-risk; adults who were immunocompetent with 1 or more chronic medical condition were categorized as At-risk; and adults without evidence of any High-risk or At-risk condition were categorized as Healthy or Low-risk [10]. Additional details on the list of conditions in each of the risk categories can be found in Pelton et al., 2019 [10].

Given the lack of risk-specific burden data, health research analysts or policy makers looking at the data presented in the "incidence rate" column in Table 1 may be compelled to make assumptions about the subgroup or subpopulation burden estimates. In fact, it is quite tempting to assume that the entire population burden is similar in magnitude to the burden for the largest subgroup (i.e., low-risk, in this case) because they essentially make up the entire population (almost 90 %), and then apply the assumed relative risks to obtain the burden for the other subgroups. Such an assumption can result in estimates that are substantially (quantitatively and qualitatively) incorrect. Consequently, the data question is: can the burden (incidence, prevalence, or average cost) for each of the risk groups be calculated using data on the overall disease burden for the entire population, risk group profile, and the assumed relative burdens/values?

Table 1
Overall outpatient NBPP incidence, relative risks and risk distribution for US adults aged 18–49 years.

Risk profile	Estimate (95 % CI)	Relative risk/ rate ratio [10] Mean (95 % CI)	Risk distribution (%) [10]
% *S. pneumoniae* [28]	11.0 (9.0,13.0)	–	–
% outpatient	69.4 (63.7,73.4)	–	–
Overall all-cause (cases per 100,000)[a]	680	–	–
	Incidence rate, per 100,000 (95 % CI)		
Low-risk	? (?,?)	1	87.3
At-risk	? (?,?)	5.8 (5.6, 6.0)	10.4
High-risk	? (?,?)	19.7 (19.1, 20.5)	2.3

[a] The overall outpatient incidence rate of pneumococcal pneumonia for those aged 18–49 was calculated by applying the proportion (11 % [28]) of *S. pneumoniae* in all-cause community-acquired pneumonia (680 [11]) and the proportion of the pneumonia cases that were outpatient (69.4 % [11]) for those aged 18–49 from Tong et al., 2018 [11] (i.e., $51.9 = 680 \times 0.694 \times 0.11$).

2.2. Methodological approach

To find the subgroup-specific incidence estimates as presented in Table 1, a simple numerical example was used to illustrate the methodology. Then, a generic symbolic formula for calculating the burden (incidence, prevalence or average cost) for n subgroups or subpopulations were derived.

2.3. Example of calculating incidence for 3 subgroups

The incidence of a disease is among the most common measures of disease burden. Incidence of a disease is defined as the number of new cases of that disease that occur in a population during a given period of time [26]. Here, the incidence of NBPP is reported per 100,000 population per year. Using the data presented in Table 1, let c_1, c_2 and c_3 be the number of cases identified from each risk group and i_1, i_2 and i_3 be the incidence of disease in each risk group (healthy/low-risk, at-risk, and high-risk, respectively) in a representative 100,000 population—subscripts represent the risk groups as follows: 1 for healthy/low-risk, 2 for at-risk, and 3 for high-risk. The overall incidence (i.e., the total number of cases in a representative 100,000 population [51.91 = 680 × 0.694 × 0.11]) is:

$$51.91 = c_1 + c_2 + c_3$$

Given the respective risk group proportions in the population (87.3 %, 10.4 %, and 2.3 %), the denominators for each risk group from a representative 100,000 population will be 87,300, 10,400 and 2300, respectively. This implies that the incidence rate (cases /100,000) for each risk group can be written respectively as:

$$i_1 = \left(\frac{c_1}{87,300}\right)*100,000,$$
$$i_2 = \left(\frac{c_2}{10,400}\right)*100,000,$$
$$i_3 = \left(\frac{c_3}{2,300}\right)*100,000$$

and

$$c_1 = i_1*0.873,\ \ c_2 = i_2*0.104,\ \ c_3 = i_3*0.023$$

Thus, the overall incidence is the sum of the proportion-weighted unknown incidence in each risk group in the population represented by the following equation:

$$0.873*i_1 + 0.104*i_2 + 0.023*i_3 = 51.91 \quad (1)$$

Incorporating the relative risk estimates among the risk groups from Table 1, the following is obtained:

$$\frac{i_2}{i_1} = 5.8,$$

which implies that.

$$i_2 = 5.8*i_1, \quad (2)$$

and

$$\frac{i_3}{i_1} = 19.7.$$

which implies that

$$i_3 = 19.7*i_1. \quad (3)$$

The three Eqs. (1), (2), and (3), with three unknowns (i_1, i_2, and i_3) can be solved by substituting Eqs. (2) and (3) into Eq. (1) and obtaining i_1 and using its value in Eqs. (2) and (3) to obtain the value of i_2 and i_3. The numeric solutions are:

$$i_1 = 26.91,$$
$$i_2 = 156.06,$$
$$i_3 = 530.06.$$

2.4. 3-subgroup symbolic solutions for incidence

Using the same symbols as above (c and i for cases and incidence rates, respectively), and for an overall incidence (I) and respective proportions (d_1, d_2 and d_3), the overall incidence can be written in the form of Eq. (1) as:

$$I = d_1 i_1 + d_2 i_2 + d_3 i_3$$

If the rate ratios for subgroups 2 and 3 to subgroup 1 are r_2 and r_3, respectively, then substituting for i_2 and i_3 (in terms of i_1 as shown below),

$$i_2 = r_2 i_1 \text{ and } i_3 = r_3 i_1$$

and solving for each subgroup's incidence gives the formulae for calculating the subgroup incidence as follows:

$$i_1 = \frac{I}{d_1 + d_2 r_2 + d_3 r_3},\ \ i_2 = \frac{I\, r_2}{d_1 + d_2 r_2 + d_3 r_3},\ \ i_3 = \frac{I\, r_3}{d_1 + d_2 r_2 + d_3 r_3}$$

By substituting the respective values in these formulae, the numeric results obtained in the preceding example can be confirmed.

2.5. 3-subgroup symbolic solutions for prevalence

A frequently used measure of disease burden is prevalence which is defined as the proportion of the population with existing cases of disease in a given location and at a particular time [26].

Using the same symbols as above (*c, r, and d* for cases, rate ratio, and proportions respectively), subgroup prevalence rate (p) and overall population size (N), the overall prevalence (P) can be written as:

$$P = \frac{c_1 + c_2 + c_3}{N} = \frac{c_1}{N} + \frac{c_2}{N} + \frac{c_3}{N}$$

and because

$$p_1 = \frac{c_1}{d_1 N},\ \ p_2 = \frac{c_2}{d_2 N},\ \ p_3 = \frac{c_3}{d_3 N}$$

Overall prevalence can be written in the same form as Eq. (1):

$$P = d_1 p_1 + d_2 p_2 + d_3 p_3$$

Again, the overall prevalence is the sum of the proportion-weighted individual subgroup prevalence rates. Substituting for p_2 and p_3 (in terms of p_1 as shown below),

$$p_2 = r_2 p_1 \text{ and } p_3 = r_3 p_1$$

and solving for each subgroup prevalence rate gives the formulae for calculating the subgroup prevalence rates as follows:

$$p_1 = \frac{P}{d_1 + d_2 r_2 + d_3 r_3},\ \ p_2 = \frac{P\, r_2}{d_1 + d_2 r_2 + d_3 r_3},\ \ p_3 = \frac{P\, r_3}{d_1 + d_2 r_2 + d_3 r_3}$$

2.6. 3-subgroup symbolic solutions for average treatment cost

Average cost is defined as the arithmetic mean of the estimated treatment cost per patient—the sum of the treatment costs for each patient dividend by the number of patients. Using the same symbols as above (*r, d, and N* for ratio, proportions, and total population size, respectively) and subgroup total cost (T), the overall average cost (A) can be written as:

$$A = \frac{T_1 + T_2 + T_3}{N} = \frac{T_1}{N} + \frac{T_2}{N} + \frac{T_3}{N}$$

and because subgroup average costs are calculated as follows:

$$a_1 = \frac{T_1}{d_1 N},\ a_2 = \frac{T_2}{d_2 N},\ a_3 = \frac{T_3}{d_3 N}$$

The overall average cost can be written in the form of Eq. (1):

$$A = d_1 a_1 + d_2 a_2 + d_3 a_3$$

As demonstrated above, the overall average cost is the sum of the proportion-weighted individual subgroup average costs. Substituting for a_2 and a_3 (in terms of a_1 as shown below),

$$a_2 = r_2 a_1 \text{ and } a_3 = r_3 a_1$$

and solving for each subgroup's average cost gives the formulae for calculating the subgroup average costs as follows:

$$a_1 = \frac{A}{d_1 + d_2 r_2 + d_3 r_3},\ a_2 = \frac{A\, r_2}{d_1 + d_2 r_2 + d_3 r_3},\ a_3 = \frac{A\, r_3}{d_1 + d_2 r_2 + d_3 r_3}$$

2.7. Generalized formulae for calculating estimates for n subgroups

A generic formula for n subgroups can be derived by extending the logic in the symbolic solutions derived for the 3-subgroup formulae presented in the preceding sections. For n subgroups, the incidence/prevalence rate or average cost m_s for any subgroup s can be calculated given information on an overall measure of disease burden and relative magnitude of the measure for the subgroups.

As demonstrated in Eq. (1) and subsequent sections, the overall measure of disease burden (incidence/prevalence rate or average cost) is the sum of the proportion-weighted individual subgroup measures m_s:

$$\sum_{s=1}^{n} d_s m_s = M. \tag{4}$$

The relative measure in subgroup s or rate ratios can be calculated as:

$$r_s = \frac{m_s}{m_1}, s = 1, 2, \ldots, n. \tag{5}$$

Equation can be solved for m_s as

$$m_s = r_s m_1, s = 1, 2, \ldots, n. \tag{6}$$

By substituting Eq. (6) into Eq. (4) and solving for m_1 yields

$$m_1 = \frac{M}{\sum_{s=1}^{n} r_s d_s},$$

which can be used with Eq. (6) to get

$$m_s = \frac{r_s M}{\sum_{j=1}^{n} r_j d_j}, s = 1, 2, \ldots, n. \tag{7}$$

2.8. Interpretation

The derived formulae indicated that each subgroup or subpopulation's burden can be calculated as the overall burden (incidence, prevalence, or average cost) adjusted by the ratio of that subgroup or subpopulation's relative burden to the sum of the proportion-weighted relative burdens of all the subgroups or subpopulations within the population—in other words, the denominator is the sum of the product of the relative burdens and their respective proportions. The formulae also demonstrate that, as expected for a particular population and corresponding subgroups or subpopulations within it, the magnitude of the calculated subgroup or subpopulation's burden compared to those of the other subgroup or subpopulation is determined by its relative morbidity. This is because, as shown in the derived formulae, the only quantity that is different in the derived formulae for calculating each subgroup or subpopulation burden is the relative burden—the overall burden and the aggregated proportion-weighted relative morbidities are the same for that population.

2.9. Estimating uncertainty

It is important to account for uncertainty in the point estimate of the sub-group specific measure of disease burden reported in equation. For example, the uncertainty in the estimate of incidence of disease should reflect the underlying uncertainty in incidence and relative risks as shown in Table 1. Assuming Gamma distributions, with parameters estimated from the confidence intervals, the mean (and uncertainty intervals) of sub-group incidence are (Supplementary Appendix).

$$i_1 = 26.91(21.9932.20),$$
$$i_2 = 156.03(127.57186.74),$$
$$i_3 = 529.95(433.30, 634.26).$$

2.10. Application

Although this report focused on three major measures of disease burden (incidence, prevalence, or average cost) for health risk subgroups, the methodological approach to the derivation of the formulae can be applied to several epidemiologic or population-level measures (such as case fatality rate, screening or testing rate, vaccine coverage rate, (sexual) contact rate, crime rate, birth rate and incarceration rate), and for different subgroups or subpopulations (such as those based on race or ethnicity, age, gender, and sexual orientation) within a large population. However, for some of the measures, imposing some mathematical constraints may be necessary to obtain realistic estimates. For instance, when calculating (sexual) contact rates, the number of contacts need to match for any pair of partners within each of the subgroups or subpopulations.

3. Discussion

Accurate (or good) data play an indispensable role at every stage (initiation, planning, executing, evaluating or monitoring, and modifying) in the lifecycle of public health interventions and research. Although good data on the entire population of interest may be available, data for some subgroups or subpopulations may not. In this report, a set of generic formulae that can be used to calculate subgroup burden or measure using the overall estimates and the reported or assumed relative burden for the three major population-level measures of disease burden (incidence, prevalence, or average cost) were derived. The formulae indicated that the subgroup or subpopulation's disease burden can be calculated as the overall burden (incidence, prevalence, or average cost) adjusted by the ratio of that subgroup or subpopulation's relative burden to the sum of the proportion-weighted relative burdens of all the subgroups or subpopulations within the population.

3.1. Strengths

There are several strengths with the methodological approach and formulae presented in this report. First, simple intuitive rationales were provided for the equations used to derive the formulae. Second, the quantitative (and qualitative) results were confirmed using manual calculation and two software packages—Microsoft Excel 365 ® (Microsoft Corporation, Redmon, WA, USA) and Mathematica® 14.1 (Wolfram Research, Champaign, IL). Third, the generic symbolic formulae presented provide flexibility in terms of the number of subgroups or subpopulations that the formulae can be used for. In other words, the formulae can be used to calculate the subgroup or subpopulation metrics for any number of subgroups or subpopulations, provided the overall metric and relative metrics are available, and the converse is also true—the overall burden can be calculated from the

subgroup or subpopulation burdens and relative burdens for any number of subgroups or subpopulations. Fourth, although the formulae were derived specifically for disease burden (incidence, prevalence, or average cost), the methodological approach can be applied to many other epidemiologic or population-level measures as listed above. Fifth, the formulae can be used to calculate the number of cases for each subgroup or subpopulation.

Finally, using these formulae provide quantitative and qualitative results that are consistent with expectation and minimizes or avoids data errors. For instance, had the incidence for the largest group (low-risk) in Table 1, been assumed to be the incidence for the overall population, the estimated incidence after applying the relative risks for all the risk groups would have almost doubled—approximately 100 % error (26.91 vs. 51.91 for low-risk, 156.06 vs. 301.08 for at-risk, and 530.06 vs. 1022.63) and the errors in absolute difference is even worse. In addition, the relative magnitude of the incidence among the risk groups and the overall population would not be consistent with expectation. This is because, being the weighted average of the incidence of its component subgroups or subpopulations, the magnitude of the overall incidence should fall within the range of those for its component subgroups or subpopulations, and not at the low-, high-end or outside that range.

The burden consists of both health and economic consequences of disease. This report used three commonly used measures— incidence, prevalence, and cost. However, there are other measures of the burden of disease such as the disability-adjusted life year (DALY) [27]. Even though this method is not illustrated with the DALY estimate, it is also applicable as the calculation of the DALY requires detailed information on the incidence of disease and prevalence of risk factors by subgroups. In addition, the formulae derived in this report can be used to estimate subgroup DALYs using the total, relative magnitudes and proportions as demonstrated in the preceding.

3.2. Limitations

The methodological approach in this report has one major limitation—the assumption(s) about relative burden. The derived formulae require that one has some information about the relative burden for each of the subgroups or subpopulations of the population. This limitation can be minimized by ensuring that the underlying assumption about the relative burdens (incidence, prevalence, or average cost) used in the formulae are based on good data or have scientifically tenable rationale (s). However, it is important to note that the lack of good data or assumption about the inputs that are required by the formulae to provide correct results does not imply that the formulae themselves are wrong. In fact, these formulae can be thought of as the correct approach to obtaining the subgroup or subpopulation estimates when the decision on the applicable assumptions about the relative burdens have been made. Consequently, it is imperative that all the data and underlying assumptions used in the formulae are based on robust and tenable scientific rationale.

3.3. Conclusion

All populations are made up of subgroups or subpopulations. As a result, the disease burden (incidence, prevalence, or average cost) of the subgroup or subpopulation are equally important at all stages of public health research or interventions—initiating, planning, executing, evaluating and modifying. However, due to lack of data, health research analysts or policy makers may make assumptions about the burden of disease among the subgroup or subpopulations that may over- or underestimate the burden for the subgroups or subpopulations within the population. Using over- or underestimated burden in economic evaluation (such as CEA or BIAs), epidemiologic or mathematical disease models can result in specious quantitative and qualitative results, which may ultimately be used to develop (or support) health policy recommendations that are likely to fail, in terms of achieving their set goals. In this report, a set of formulae were derived for the three major population-level measures of disease burden (incidence, prevalence, or average cost) that can be used to calculate subgroup burden using the overall burden and the reported or assumed relative burden estimates.

The derived formulae indicated that each subgroup or subpopulation's burden can be calculated as the overall burden adjusted by the ratio of that subgroup or subpopulation's relative burden to the sum of the proportion-weighted relative burdens of the subgroups or subpopulations. These formulae can help to avoid—or at the very least minimize—potential quantitative and qualitative errors in subgroup or subpopulation disease burden that are used for health research, interventions and/or policy analyses or deliberations. In addition, because they calculate disease burden among subgroups or subpopulations, they can be useful when planning and setting the appropriate (and achievable) quantitative and qualitative health equity goals as outlined in Healthy People 2030 [2].

Authors' contribution

KOE, AD and EHE conceptualized and determined the scope of the study. AD assembled the data, conducted the initial data synthesis, and posed the data question. KOE and EHE performed the mathematical derivations to arrive at the formulae. KOE prepared the first draft of the manuscript. KOE, AD and EHE reviewed the manuscript for intellectual content. All authors approved submission of the manuscript for publication.

CRediT authorship contribution statement

Kwame Owusu-Edusei: Writing – review & editing, Writing – original draft, Validation, Methodology, Conceptualization. **Arijita Deb:** Writing – review & editing, Validation, Data curation, Conceptualization. **Elamin H. Elbasha:** Writing – review & editing, Validation, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Kwame Owusu-Edusei reports a relationship with Merck & Co., Inc. that includes: employment. Elamin Elbasha reports a relationship with Merck & Co., Inc. that includes: employment. Arijita Deb reports a relationship with GSK, that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dialog.2025.100216.

References

[1] Cong B, Dighero I, Zhang T, et al. Understanding the age spectrum of respiratory syncytial virus associated hospitalisation and mortality burden based on statistical modelling methods: a systematic analysis. BMC Med 2023;21(1):224.
[2] US Department of Health and Human Services. Healthy People 2030: Health Equity in Healthy People 2030 [cited 2024 August 20]; Available from, https://health.gov/healthypeople/priority-areas/health-equity-healthy-people-2030; 2024.
[3] Gierke Ryan, Wodi Patricia, Kobayashi Miwako. Pneumococcal disease. In: The pink book: epidemiology and prevention of vaccine-preventable diseases. Atlanta, GA: National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention; 2021.
[4] Said MA, Johnson HL, Nonyane BA, et al. Estimating the burden of pneumococcal pneumonia among adults: a systematic review and meta-analysis of diagnostic techniques. PloS One 2013;8(4):e60273.

[5] Kobayashi M. Evidence to recommendation framework: risk-based us of 15-valent and 20-valent pneumococcal conjugate vaccines in adults: presented to the ACIP September 29, 2021. Centers for Disease Control and Prevention; 2021.
[6] Centers for Disease Control and Prevention. Active bacterial core surveillance report, emerging infections program network, *Streptococcus pneumoniae*, 2020. Atlanta, GA: Centers for Disease Control and Prevention; 2020.
[7] Centers for Disease Control and Prevention. Active bacterial core surveillance report, emerging infections program network, *Streptococcus pneumoniae*, 2019. Atlanta, GA: Centers for Disease Control and Prevention; 2019.
[8] European Centre for Disease Prevention and Control. Surveillance Atlas of infectious diseases. Solna, Sweden: European Centre for Disease Prevention and Control; 2022.
[9] Isturiz R, Grant L, Gray S, et al. Expanded analysis of 20 pneumococcal serotypes associated with radiographically confirmed community-acquired pneumonia in hospitalized US adults. Clin Infect Dis 2021;73(7):1216–22.
[10] Pelton SI, Bornheimer R, Doroff R, et al. Decline in pneumococcal Disease attenuated in older adults and those with comorbidities following universal childhood PCV13 immunization. Clin Infect Dis 2019;68(11):1831–8.
[11] Tong S, Amand C, Kieffer A, Kyaw MH. Trends in healthcare utilization and costs associated with pneumonia in the United States during 2008-2014. BMC Health Serv Res 2018;18(1):715.
[12] Altawalbeh SM, Wateska AR, Nowalk MP, et al. Cost-effectiveness of an in-development adult-formulated 21-valent pneumococcal conjugate vaccine in US adults aged 50 years or older. Vaccine 2024;42(12):3024–32.
[13] Altawalbeh SM, Wateska AR, Nowalk MP, et al. Societal cost of racial pneumococcal Disease disparities in US adults aged 50 years or older. Appl Health Econ Health Policy 2024;22(1):61–71.
[14] Altawalbeh SM, Wateska AR, Nowalk MP, et al. Pneumococcal vaccination strategies in 50-year-olds to decrease racial disparities: a US societal perspective cost-effectiveness analysis. Value Health 2024;27(6):721–9.
[15] Leidner A. Summary of three economic analyses on the use of 21-valent pneumococcal conjugate vaccine (PCV21) among adults in the United States. In: Advisory committee on immunization; 2024. Atlanta, GA, USA.
[16] Stoecker C. Economic assessment of PCV21 in U.S. adults. Ctr Dis Control Prev 2024.
[17] Wateska AR, Nowalk MP, Lin CJ, et al. Cost-effectiveness of an in-development adult-formulated pneumococcal vaccine in older US adults. Vaccine 2023;41(30):4431–7.
[18] Stoecker C. Economic assessment of PCV20 for adults vaccinated with PCV13: presented at ACIP, October 19, 2022. Ctr Dis Control Prev 2022.
[19] Wateska AR, Patricia Nowalk M, Lin CJ, et al. Cost-effectiveness of revised US pneumococcal vaccination recommendations in underserved minority adults < 65-years-old. Vaccine 2022;40(50):7312–20.
[20] Stoecker C. Economic assessment of PCV15 & PCV20: presentation to the ACIP June 25, 2021. Ctr Dis Control Prev 2021.
[21] Leidner AJ. Summary of three economic models Assesssing pneumococcal vaccines in US adults: presentation to the ACIP September 9, 2021. Centers for Disease Control and Prevention; 2021.
[22] Leidner A, Bletnitsky S. Summary of three economic analyses on the use of PCVs among 50-64 year old adults in the United States. In: Advisory committee on immunization; 2024. Atlanta, GA, USA.
[23] Yi Z, Johnson KD, Owusu-Edusei K. Lifetime health and economic burden of invasive pneumococcal diseases attributable to V116 serotypes among adults in the United States. Infect Dis Ther 2024;13(7):1501–14.
[24] Yi Z, Owusu-Edusei K, Elbasha E. Cost-effectiveness analysis of the use of V116, a 21-valent pneumococcal conjugate vaccine, in vaccine-naive adults aged ≥ 65 years in the United States. Infect Dis Ther 2024;13(12):2597–615.
[25] Mueller PP, Tajima A, Cassell K, et al. Health and economic impact of the 21-valent pneumococcal conjugate vaccine (V116) for adults in Japan: a delta price approach. J Med Econ 2025;28(1):136–45.
[26] Noordzij M, Dekker FW, Zoccali C, Jager KJ. Measures of disease frequency: prevalence and incidence. Nephron Clin Pract 2010;115(1):c17–20.
[27] Collaborators GBDV. Five insights from the global burden of disease study 2019. Lancet 2020;396(10258):1135–59.
[28] World Health Organisation (WHO). Strategic advisory group of experts on immunization. Geneva, Switzerland: World Health Organization; 2020.