An easy-to-use computer program for standardisation methods of population morbidity data

Suan Mei Ong, Wan Nor Arifin and Najib Majdi Yaacob

Unit of Biostatistics and Research Methodology,
School of Medical Sciences,
Health Campus,
Universiti Sains Malaysia,
16150 Kubang Kerian, Kelantan, Malaysia
Email: suanmei.ong@gmail.com
Email: wnarifin@usm.my
Email: najibmy@usm.my

Nyi Nyi Naing*

Institute for Community (Health) Development (i-CODE),
Universiti Sultan Zainal Abidin,
Block E, Level 1, Gong Badak Campus,
21300 Kuala Nerus, Terengganu, Malaysia
Email: syedhatim@unisza.edu.my
*Corresponding author

Abstract: Standardisation is an essential procedure to eliminate the effect of confounding when comparisons between populations are carried out, where a standard population is used as a reference. There are two methods of standardisation, i.e., direct and indirect standardisation. Standardisation is commonly used in epidemiology studies especially when the morbidity or/and mortality rates of a disease are studied. A computer program (StdAn) which aims to simplify the process of standardisation of population morbidity data was developed with Microsoft Visual Studio 2010 Express software, using C++/CLI (C++ on common language infrastructure) as the programming language. StdAn program provides a graphically easy-to-use interface for the implementation of standardisation analysis. It is simple, practical and easy to interpret.

Keywords: standardisation analysis; direct standardisation; indirect standardisation; standardised morbidity rate; computer program.

Reference to this paper should be made as follows: Ong, S.M., Arifin, W.N., Yaacob, N.M. and Naing, N.N. (2020) 'An easy-to-use computer program for standardisation methods of population morbidity data', *Int. J. Computing Science and Mathematics*, Vol. 11, No. 1, pp.40–53.

Biographical notes: Suan Mei Ong is a PhD student majoring in Biostatistics at the School of Medical Sciences, Universiti Sains Malaysia. Her current research interests include statistics in medical research, machine learning, statistical modelling and computational mathematics.

Wan Nor Arifin is a Senior Lecturer at the School of Medical Sciences, Universiti Sains Malaysia. His current research interests include factor analysis, questionnaire validation and biostatistics.

Najib Majdi Yaacob is a Senior Lecturer at the School of Medical Sciences, Universiti Sains Malaysia. His current research interests include epidemiology, biostatistics and public health.

Nyi Nyi Naing is a Professor at the Institute for Community (Health) Development (i-CODE), Universiti Sultan Zainal Abidin, Block E, Level 1, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia. His current research interests include TB-HIV coinfection, smoking related research and statistical modelling.

This paper is a revised and expanded version of a paper entitled 'A computer program for easy use of standardisation methods of population morbidity data' presented at International Conference on Computational Methods in Engineering and Health Sciences (ICCMEH-2015), Universiti Putra Malaysia, Malaysia, 19–20 December 2015.

1 Introduction

Many studies involve comparisons between two or more populations. In health, researchers were interested to compare the patients' morbidity (McLean et al., 2017; Wojtyła et al., 2016; Videnović et al., 2015) or/and mortality rates (Perazzo et al., 2017; Vanthomme et al., 2015; Lee et al., 2015) of diseases between populations. In socio-economic studies, inequality of children disability (Raistenskis et al., 2015), alcohol-related studies (Thompson et al., 2017; McCartney et al., 2016) and dietary patterns (Ferraro et al., 2016; Orlich et al., 2014) were studied between communities. In insurance, the prevalence of a disease was studied by using the insurees' claim data (Hadji et al., 2013) for various locations. In education, tertiary enrolment rates could be compared between populations and many more other examples from various sectors.

As the populations involved in the comparisons are usually diverse in composition due to sociodemographic (such as age, sex, ethnicity), geographic and other factors, therefore comparisons between populations by using crude rates are not appropriate and can be misleading as they are calculated based on the populations under study alone. Comparison by using crude rate is only applicable and meaningful if the subgroups of the populations are alike in other factors which are associated with the topic of interest or else these factors may become the confounders in the study. For example, the crude mortality rates from coronary heart disease in population A, which has older people, should be higher than younger population B. In order to remove the effect of age in this instance, comparison can be done on the mortality rates of these populations in the same age subgroups, where a standard population is used as a reference. This process is called standardisation. In this example, the confounding variable is age, therefore it is referred as age standardisation and the standardised rate calculated is termed as age-standardised mortality rates. If sex standardisation is applied simultaneously, it leads to age/sex standardisation (Woodward, 2013).

There are two methods commonly used in calculating standardised rates, i.e., direct and indirect standardisation. Direct standardisation can be employed when factor-specific

rates of the populations under study are known, else indirect standardisation is applied, where factor-specific rates of standard population is involved in the analysis. The standard population selected is usually a superset of the populations involved or one of the study populations (Woodward, 2013). Therefore, if we would like to compare tuberculosis morbidity rates between the states in Malaysia, we could either use the entire population of Malaysia, a few states or one of the states as the standard population.

Standardisation analysis (StdAn) is commonly carried out by using spreadsheet program such as Microsoft Excel. Another existing computer program which has the function of standardisation is WinPepi program, which is a computer program that comprises various statistical programs developed for epidemiologists (Abramson, 2011).

The objective of this study was to develop a computer program (StdAn) that could simplify the process of standardisation of population morbidity data. The layout of StdAn program is an adaptation of the illustrations in Naing (2000). By using StdAn program, users can select to perform either direct or/and indirect standardisation on their population morbidity data. StdAn program enables the StdAn to be carried out with cheaper cost and shorter time.

2 Methodology

For illustration purposes, age factor and morbidity rate are used to explain the procedure of direct and indirect standardisation.

2.1 Direct standardisation

When age-specific morbidity rates of the populations under study are known, direct standardisation can be applied. The procedure for direct standardisation is elaborated step-by-step in Table 1.

Table 1 Direct standardisation procedure

No.	*Activity*	
1	For each age subgroup, calculate the age-specific morbidity rate of population (1 and 2) by dividing the number of morbidity to the population at risk respectively.	
	Age-specific morbidity rate = Number of morbidity/population at risk	1
2	Select the standard population.	
3	For each age subgroup, calculate the expected morbidity of population (1 and 2) by multiplying the age-specific morbidity rate of population (1 and 2) to the respective subgroup of the standard population.	
	Expected morbidity = Age-specific morbidity rate × standard population	2
4	Sum up the expected morbidity for all the age subgroups of population (1 and 2) respectively.	
5	Calculate the age-standardised morbidity rates of population (1 and 2) by dividing total expected morbidity of population (1 and 2) to the total standard population respectively.	
	Age-standardised morbidity rate = Total expected morbidity/total standard population	3
6	Conclude by comparing the age-standardised morbidity rates of population (1 and 2) to find out which population has the higher rate of morbidity.	

2.2 Indirect standardisation

When age-specific morbidity rates of the standard population are known, indirect standardisation can be employed. The procedure for indirect standardisation is explained step-by-step in Table 2.

Table 2 Indirect standardisation procedure

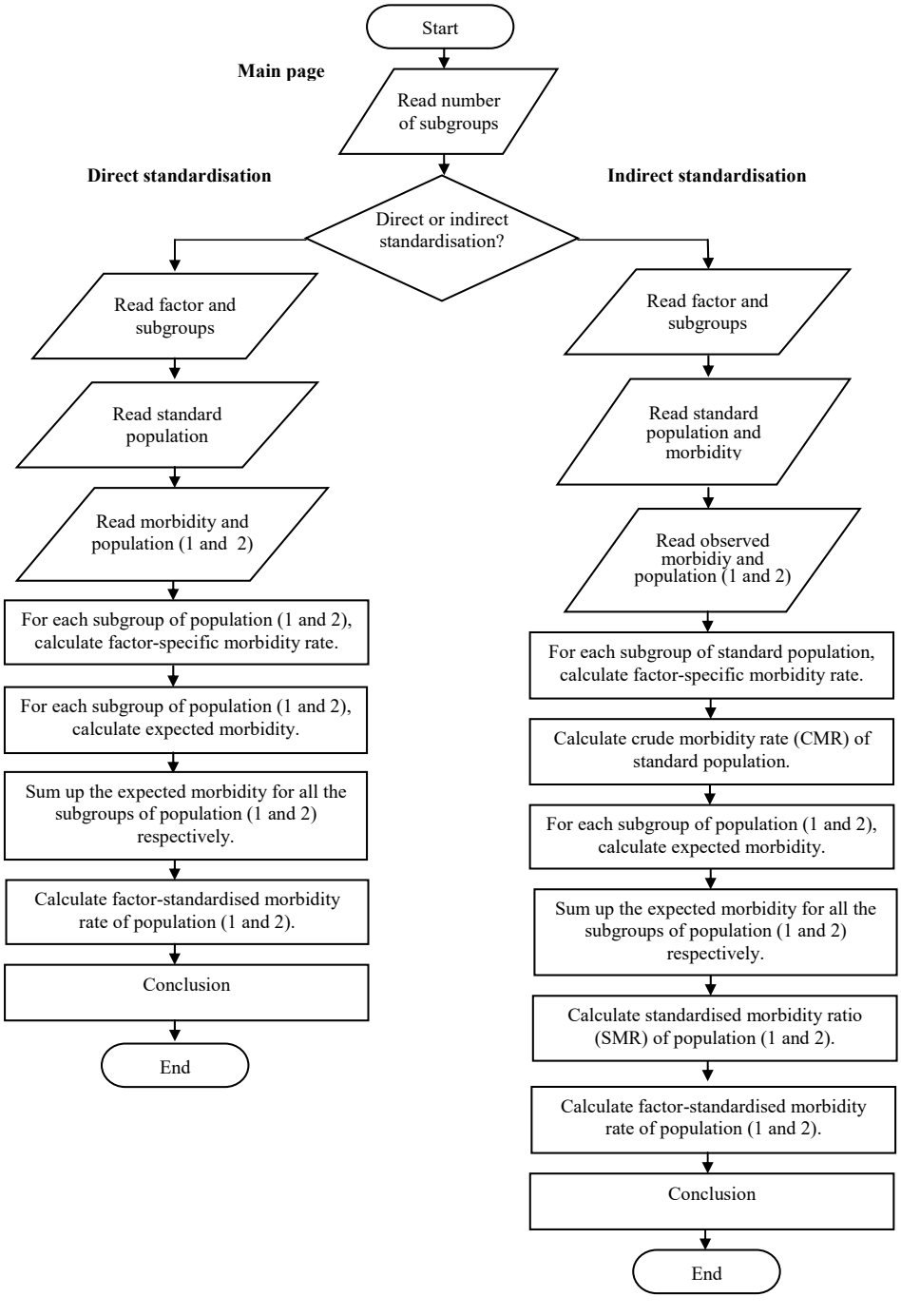
No.	*Activity*	
1	Select the standard population. For each age subgroup, calculate the age-specific morbidity rate of standard population by dividing the number of morbidity to the population at risk.	
	Age-specific morbidity rate = Number of morbidity/population at risk	4
2	Calculate crude morbidity rate of standard population by dividing total morbidity to total standard population.	
	Crude morbidity rate (CMR) = Total morbidity/total standard population	5
3	Obtain the observed morbidity of population (1 and 2).	
4	For each age subgroup, calculate the expected morbidity of population (1 and 2) by multiplying the age-specific morbidity rate of standard population to the population at risk of population (1 and 2) respectively.	
	Expected morbidity = Age-specific morbidity rate (standard × population at risk)	6
5	Sum up the expected morbidity for all the age subgroups of population (1 and 2) respectively.	
6	Calculate the standardised morbidity ratio (SMR) of population (1 and 2) by dividing the observed morbidity of population (1 and 2) to the total expected morbidity of population (1 and 2) respectively.	
	Standardised morbidity ratio (SMR) = Observed morbidity/total expected morbidity	7
7	Calculate the age-standardised morbidity rate of population (1 and 2) by multiplying SMR of population (1 and 2) to CMR of standard population respectively.	
	Age-standardised morbidity rate = SMR × CMR (standard)	8
8	Compare the age-standardised morbidity rates of population (1 and 2) and conclude findings.	

3 The computer program

A computer program for StdAn was developed with Microsoft Visual Studio 2010 Express software, using C++/CLI (C++ on common language infrastructure) as the programming language. The Windows Forms application feature in the integrated development environment was used to create forms to display graphical information of controls for data input and output (Microsoft Corporation, 2015).

StdAn program provides a graphically easy-to-use interface for the implementation of StdAn. Three forms are available, i.e., main page, direct standardisation and indirect standardisation forms. At the main page form, user can select the method of standardisation and number of subgroups (between two to five) involved in the analysis; whereas at direct and indirect standardisation forms, execution of the direct and indirect standardisation analyses can be carried out respectively. Figure 1 displays the flowchart of StdAn in StdAn program.

Figure 1 Flowchart of direct and indirect standardisation methods



3.1 The source code

The source code of buttons in StdAn program is displayed below.

3.1.1 Exit, close and clear buttons

For the execution of Exit, Close and Clear buttons, the source code is as follows:

a To exit from StdAn program: Application::Exit( );

b To close a form: Close( );

c To clear the value of a variable, for instance, Factor A: Factor A → Text = " ";

3.1.2 Calculate button

The tasks that are performed by clicking calculate button are:

a Standard population:
- Calculate total standard population

b Population 1 and 2:
- Calculate
  1 total population
  2 total morbidity
  3 morbidity rates
  4 crude morbidity rate
  5 expected morbidity
  6 factor-standardised morbidity rate (direct and indirect standardisation).

The source code for the execution of calculate button is shown below, where factor with two subgroups is used here. The comments are written after double slash (//) to explain the code:

a Standard population: calculate total standard population

```
Double sumpop = 0, pop1 = 0, pop2 = 0;                    // Define population variables as double
pop1 = Convert::ToDouble(text2Stdpop1 → Text);            // Convert subgroup 1 from text to double
pop2 = Convert::ToDouble(text2Stdpop2 → Text);            // Convert subgroup 2 from text to double
sumpop = pop1 + pop2;                                     // Sum up subgroups 1 and 2
text2TotalStdpop → Text = Convert::ToString(sumpop);      // Convert sum from double to text
```

b The source codes for population 1 only is shown as population 2 has similar code:

1 Calculate total population 1

```
Double sumpop1 = 0, pop1_1 = 0, pop1_2 = 0;          // Define population variables as double
pop1_1 = Convert::ToDouble(text2Pop1_1 → Text);      // Convert population from text to double
pop1_2 = Convert::ToDouble(text2Pop1_2 → Text);      // Convert population from text to double
sumpop1 = pop1_1 + pop1_2;                           // Sum up population of subgroups 1 and 2
text2TotalPop1 → Text = Convert::ToString(sumpop1);  // Convert sum from double to text
```

2 Calculate total morbidity of population 1

```
Double summor1 = 0, mor1_1 = 0, mor1_2 = 0;          // Define morbidity variables as double
mor1_1 = Convert::ToDouble(text2Mor1_1 → Text);      // Convert morbidity from text to double
mor1_2 = Convert::ToDouble(text2Mor1_2 → Text);      // Convert morbidity from text to double
summor1 = mor1_1 + mor1_2;                           // Sum up morbidity of subgroups 1 and 2
text2TotalMor1 → Text = Convert::ToString(summor1);  // Convert sum from double to text
```

3 Calculate morbidity rates of population 1

```
Double MR1_1 = 0, MR1_2 = 0;                         // Define morbidity rate variables as double
MR1_1 = mor1_1/pop1_1;                               // Calculate morbidity rate of subgroup 1
MR1_2 = mor1_2/pop1_2;                               // Calculate morbidity rate of subgroup 2
text2MR1_1 → Text = MR1_1.ToString("F3");            // Convert MR1_1 from double to text
text2MR1_2 → Text = MR1_2.ToString("F3");            // Convert MR1_2 from double to text
```

4 Calculate crude morbidity rate of population 1

```
Double CMR1 = 0;                                     // Define crude morbidity rate variable as double
CMR1 = summor1/sumpop1;                              // Calculate crude morbidity rate
```

```
text2CMR1 → Text = CMR1.ToString("F3");          // Convert CMR1 from double to text
```

5 Calculate expected morbidity of population 1

```
Double sumEM1 = 0, EM1_1 = 0, EM1_2 = 0;         // Define expected morbidity variables as double
EM1_1 = pop1*MR1_1;                              // Calculate expected morbidity
EM1_2 = pop2*MR1_2;                              // Calculate expected morbidity
sumEM1 = EM1_1 + EM1_2;                          // Sum up expected morbidity of subgroups 1 and 2
text2ExMor1_1 → Text = EM1_1.ToString("F3");     // Convert EM1_1 from double to text
text2ExMor1_2 → Text = EM1_2.ToString("F3");     // Convert EM1_2 from double to text
text2TotalExMor1 → Text = sumEM1.ToString("F3"); // Convert sumEM1 from double to text
```

6 Calculate factor-standardised morbidity rate of population 1

a Direct standardisation

```
Double AMR1 = 0;                                 // Define standardised morbidity rate as double
AMR1 = sumEM1/sumpop;                            // Calculate factor-standardised morbidity rate
text2AMR1 → Text = AMR1.ToString("F3");          // Convert AMR1 from double to text
```

b Indirect standardisation:

```
//Calculate standardised morbidity ratio (SMR)
Double obs1 = 0, SMR1 = 0;                       // Define variables as double
obs1 = Convert::ToDouble(txt2ObsMor1 → Text);    // Convert observed morbidity fom text to double
SMR1 = obs1/sumEM1;                              // Calculate (SMR)
txt2SMR1 → Text = SMR1.ToString("F3");           // Convert SMR1 from double to text

//Calculate factor-standardised morbidity rate
```

```
Double AMR1 = 0;                          // Define standardised morbidity rate as double
AMR1 = SMR1*StdCMR;                       // Calculate factor-standardised morbidity rate
txt2AMR1 → Text = AMR1.ToString("F3");    // Convert AMR1 from double to text
```

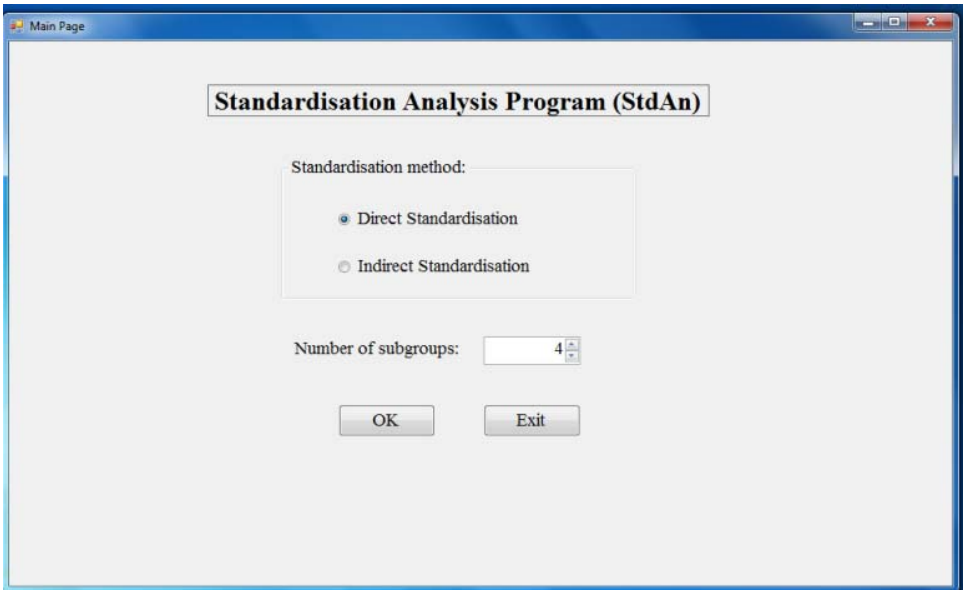
4 Validation test

Validation test was done by replicating the two examples as tabulated in Naing (2000); example 1 and example 2 are for direct and indirect standardisation respectively. Results from StdAn program were compared to the results in Naing (2000) to verify if there was any discrepancy.

4.1 Example 1 (direct standardisation)

The data provided in Example 1 of Naing (2000) are shown in Table 3. At main page form (Figure 2) of StdAn program, selected 'direct standardisation' as the standardisation method and number of subgroups as '4', then clicked 'OK' button. Program diverted to direct standardisation form (Figure 3). Data were keyed in the text boxes, followed by clicking on the 'calculate' button.

Figure 2 Main page form (see online version for colours)



The results indicated that the age-standardised morbidity rates of population 1 and population 2 were 23.262 and 22.727 per 1,000 populations respectively. Therefore, population 1 had higher risk of morbidity than population 2. Results obtained from StdAn program were the same as in Naing (2000) when numbers were rounded to one decimal place (see Table 5).

Figure 3 Direct standardisation form (see online version for colours)

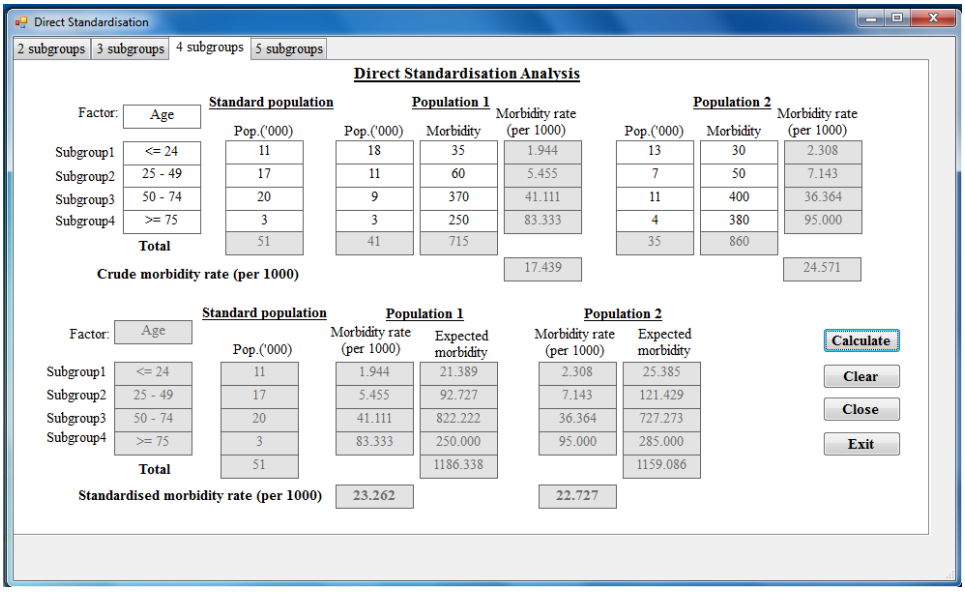


Table 3 Data given in example 1

Age group (years)	*Standard population*	*Population 1*		*Population 2*	
	Population ('000)	*Population ('000)*	*Morbidity*	*Population ('000)*	*Morbidity*
≤ 24	11	18	35	13	30
25–49	17	11	60	7	50
50–74	20	9	370	11	400
≥ 75	3	3	250	4	380
Total	*51*	*41*	*715*	*35*	*860*

Source: Naing (2000)

4.2 Example 2 (indirect standardisation)

The data provided in example 2 of Naing (2000) are shown in Table 4. At main page form (Figure 2) of StdAn program, selected 'indirect standardisation' as the standardisation method and number of subgroups as '4', then pressed 'OK' button. Program diverted to indirect standardisation form (Figure 4). Data were keyed in the text boxes, followed by clicking on the 'calculate' button.

Figure 4 Indirect standardisation form (see online version for colours)

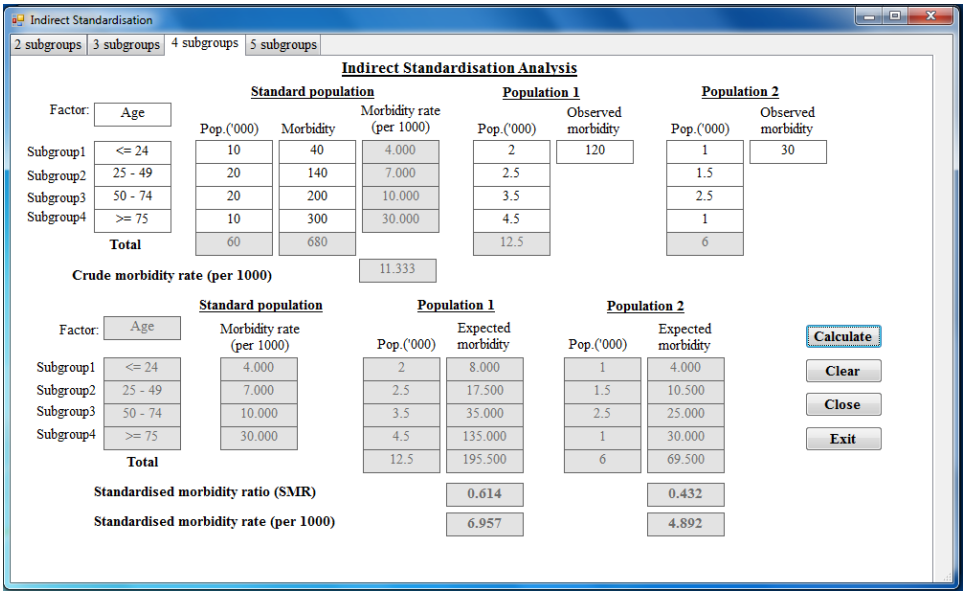


Table 4 Data given in example 2

Age group (years)	*Standard population*			*Population 1*	*Population 2*
	Population ('000)[a]	*Morbidity*[b]	*Morbidity rate (per 1,000)*	*Population ('000)*	*Population ('000)*
≤ 24	10	40	4	2	1
25–49	20	140	7	2.5	1.5
50–74	20	200	10	3.5	2.5
≥ 75	10	300	30	4.5	1
Total	*60*	*680*		*12.5*	*6*
Observed morbidity				*120*	*30*

Note: [a, b]Population and morbidity of each age subgroup in standard population were not give in Naing (2000), they were assigned by author in order to obtain the given morbidity rates.

Source: Naing (2000)

The SMR generated from StdAn program were 0.614 and 0.432 for population 1 and population 2 respectively. The results were the same as in Naing (2000) when the numbers were rounded to one decimal place (see Table 5). As the calculation of age-standardised morbidity rates in example 2 were not given in Naing (2000), we verified the accuracy of the results based on manual calculation (see 8 in Table 2).

The age-standardised morbidity rates of population 1 and population 2 obtained from StdAn program were 6.957 and 4.892 per 1,000 populations respectively. Therefore, population 1 had higher risk of morbidity as compared to population 2.

The validation test results confirmed that the StdAn program is able to perform direct and indirect standardisation accurately.

Table 5 Comparison of results in Naing (2000) and StdAn program

No.	*Analysis*	*Population*	*Naing (2000)*	*StdAn program*
1	Direct standardisation			
1	Age-standardised morbidity rate	1	23.3	23.262
	(per 1,000 population)	2	22.7	22.727
2	Indirect standardisation			
1	Standardised morbidity ratio (SMR)	1	0.61	0.614
		2	0.43	0.432
2	Age-standardised morbidity rate	1	N/A[a]	6.957
	(per 1,000 population)	2	N/A[b]	4.892

Note: [a, b]Unable to calculate age-standardised morbidity rates due to unavailability of crude morbidity rate for standard population.

5 Discussion

StdAn program is an easy-to-use graphical computer program for the implementation of StdAn on population morbidity data. The program's executable (.exe) file can be downloaded easily and free of charge (USM, 2017) or you may send your request via e-mail to suanmei.ong@gmail.com. The program can run independently without having to download the Microsoft Visual C++ 2010 Express software.

StdAn program allows users to choose to perform either direct or/and indirect standardisation methods. It enables the comparison of factor-standardised morbidity rates between two populations at each execution and the results are displayed in a single layout. Even though the column labels in StdAn program indicate that the StdAn is meant for morbidity rate, but in practice, it can also be used for other studies such as mortality rate, hospitalisation rate, etc.

Spreadsheet program such as Microsoft Excel is commonly used to perform StdAn. However, each user has to design the format of the spreadsheet individually which is time consuming and prone to human errors. Another existing computer program which has the function of standardisation is WinPepi program. In WinPepi, the StdAn is done for a single population at a time and comparison of standardised rates between populations can be done manually after that. Whereas, StdAn program allows comparison of factor-specific morbidity rate between two populations to be carried out in a single execution, moreover, users are also notified of the population that has the higher risk in morbidity.

There are several limitations to the StdAn program. One of them is the number of factor subgroups can be selected between two and five only. Besides that, it can only perform comparison between two populations at each execution and the output generated is unable to save nor print. Therefore, future program enhancement is needed to create more than five factor subgroups dynamically at run time, comparison of more than two populations at each execution and, options to save and print the output generated.

6 Conclusions

In epidemiological research, standardisation is an important process to remove the effect of confounding when comparisons between populations are carried out. StdAn program is a useful tool which has managed to simplify the StdAn on population morbidity data. It is user friendly, practical and easy to interpret. Therefore, StdAn program enables the comparison of factor-standardised morbidity rates between populations to be done efficiently with cheaper cost and shorter time. It will benefit researchers who are required to apply StdAn in their studies.

Acknowledgements

This study was supported by the Universiti Sains Malaysia Short Term Research Grant Scheme (Grant number: 304/PPSP/61313141).

References

Abramson, J.H. (2011) 'WINPEPI updated: computer programs for epidemiologists and their teaching potential', *Epidemiologic Perspectives and Innovations*, Vol. 8, No. 1, pp.1–9.

Ferraro, P.M., Curhan, G.C., Gambaro, G. and Taylor, E.N. (2016) 'Total, dietary and supplemental vitamin C intake and risk of incident kidney stones', *American Journal of Kidney Diseases*, Vol. 67, No. 3, pp.400–407.

Hadji, P., Klein, S., Gothe, H., Häussler, B., Kless, T., Schmidt, T., Steinle, T., Verheyen, F. and Linder, R. (2013) 'The epidemiology of osteoporosis – bone evaluation study (BEST): an analysis of routine health insurance data', *Deutsches Ärzteblatt International*, Vol. 110, No. 4, pp.52–57.

Lee, S.W., Kim, H.C., Lee, H.S. and Suh, I. (2015) 'Thirty-year trends in mortality from cardiovascular diseases in Korea', *Korean Circulation Journal*, Vol. 45, No. 3, pp.202–209.

McCartney, G., Bouttell, J., Craig, N., Craig, P., Graham, L., Lakha, F., Lewsey, J., McAdams, R., MacPherson, M. and Minton, J. (2016) 'Explaining trends in alcohol-related harms in Scotland, 1991–2011 (I): the role of incomes, effects of socio-economic and political adversity and demographic change', *Public health*, March, Vol. 132, pp.13–23.

McLean, G., Hindle, J.V., Guthrie, B. and Mercer, S.W. (2017) 'Co-morbidity and polypharmacy in Parkinson's disease: insights from a large Scottish primary care database', *BMC Neurology*, Vol. 17, No. 1, p.126.

Microsoft Corporation. (2015) *Microsoft Visual Studio 2010 Windows Forms Overview*, World Wide Web [online] https://msdn.microsoft.com/en-us/library/vstudio/8bxxy49h(v=vs.100).aspx (accessed 1 September 2015).

Naing, N.N. (2000) 'Easy way to learn standardization: direct and indirect methods', *The Malaysian Journal of Medical Sciences*, Vol. 7, No. 1, pp.10–15.

Orlich, M.J., Jaceldo-Siegl, K., Sabaté, J., Fan, J., Singh, P.N. and Fraser, G.E. (2014) 'Patterns of food consumption among vegetarians and non-vegetarians', *British Journal of Nutrition*, Vol. 112, No. 10, pp.1644–1653.

Perazzo, H., Pacheco, A.G., De Boni, R., Luz, P.M., Fittipaldi, J., Cardoso, S.W., Grinsztejn, B. and Veloso, V.G. (2017) 'Age-standardized mortality rates related to cirrhosis in brazil from 2000 to 2012: a nationwide analysis', *Annals of Hepatology: Official Journal of the Mexican Association of Hepatology*, Vol. 16, No. 2, pp.269–278.

Raistenskis, J., Skvarciany, Z. and Gurevičius, R. (2015) 'Regional inequalities of children disability incidence in Lithuania', *Health Sciences*, Vol. 25, Nos. 1–98, pp.33–38.

Thompson, A., Wright, A.K., Ashcroft, D.M., Van Staa, T.P. and Pirmohamed, M. (2017) 'Epidemiology of alcohol dependence in UK primary care: results from a large observational study using the clinical practice research datalink', *PloS one*, Vol. 12, No. 3, p.e0174818.

Universiti Sains Malaysia (USM) (2017) *World Wide Web* [online] http://www.medic.usm.my/biostat/.

Vanthomme, K., Vandenheede, H., Hagedoorn, P., Deboosere, P. and Gadeyne, S. (2015) 'Trends in site-and sex-specific cancer mortality between 1979 and 2010 in Belgium compared with Europe using WHO data', *Journal of Public Health*, Vol. 38, No. 2, pp.e68–e76.

Videnović, G., Miljuš, D., Ilić, D., Krasić, D. and Živković, S. (2015) 'Nonmelanoma skin cancer in the population of the city of Belgrade in the period 1999-2011', *Srpski arhiv za celokupno lekarstvo*, Vol. 143, Nos. 5–6, pp.290–295.

Wojtyła, C., Słabuszewska-Jóźwiak, A., Janik-Koncewicz, K. and Zatoński, W.A. (2016) 'New challenges for cervical cancer', *Journal of Health Inequalities*, Vol. 2, No. 1, pp.77–88.

Woodward, M. (2013) *Epidemiology Study Design and Data Analysis*, 3rd ed., chapter 4, CRC Press, Boca Raton.