**SUPPLEMENTARY MATERIAL \*\* - INCIDENCE OVER TIME**

Study selection

The primary source for papers was all papers used to provide incidence numbers for children and adolescents current Atlas and the nine previous Atlases (the first being in 2000). \*\*\* studies were identified. We also sourced all relevant papers mentioned in the identified papers, as well as all the reviews we could find.

Papers were included in the analysis if they had annual, or start and end year data, for at least five years, or if another study with the same methodology, age-group, and catchment area also had data giving a combined ≥ 5-year (y) duration.

Only paediatric data was included. Studies generally had data <15 y, however some were <20 y, and a couple <17 y, <18 y or <19 y. If <15 y and <20 data was stated in the same paper, we used <20.

Exclusions due to overlapping geographies: \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Graham to write more here \*\*\*\*\*\*\*\*\*\*\*\*\*\*

Some quirks at present that we need to decide what to do with, with some overlapping-geography studies included in the analysis:

1. **Finland** has both
2. Merged whole-country data 1953-2011 from Somersalo 1955 (cited in Ruannen (1982), Tuomilehto (1999) and Harjutsalo (2013) data, and
3. The Parvianen (2020) study which has data from 2003-2018, which are slightly different to a). Jaakko Toumilehto thinks that the data in method a) is more accurate as it is from insulin prescriptions. b) is from the national registry which may not be fully complete.

*I suggest deleting b)*

1. **Sweden** has both
2. the EURODIAB study 1989-2013 which covers Stockholm, and
3. merged data for the whole country for 1978-2007 from the Nystrom (1990), Berhan (2011) and Samuelsson (2019) studies

*I suggest deleting a), as b) has a longer period and we have few studies going back that far*

1. **Australia** has
2. Whole-country data from 2000-2017 from Haynes (2015 and 2020)
3. New South Wales data from 1990-2002 from Craig (2000) and Taplin (2005)
4. Western Australia data from 1985-2016 from Haynes (2018)

*I suggest deleting a), and using b) and c) as they are independent*

1. **USA** has
2. Colorado data from 1978-88 (one data point only) and 2002-04 (one data point only) from Vehik (2007)
3. And then Colorado is included in the SEARCH study data from 2002-15

*I suggest using both and make a note in the relevant Table*

1. **Romania** has
2. Whole-country <18-year data from 1996-2015 (Vlad (2018))
3. Bucharest <15-year data from 1992-2013 (EURODIAB)

*I suggest using b) EURODIAB as <15 and longer period?*

1. **Croatia** has:
2. Whole-country data from 1997-2012 from Stipancic (2008) and Rojnic Putarek (2014)
3. Zagreb data from 1995-2013 from EURODIAB

*I suggest deleting a) as a shorter period and ascertainment is not stated*

1. **Republic of Korea** (South Korea) has
2. Whole-country data from 1995-2016 from Kim (2016) and Chae (2020)
3. Seoul data from 1985-1998 from Ko (1994) and DIAMOND (2006)

*I suggest using both with a note, like Colorado*

If a study stated data for an odd number of years – e.g. the EURODIAB study that has 5-year intervals e.g. 1999-2003, the midpoint year was used – so 2001. If the periods used were an even number of years – e.g. 1987-1990, then we plotted the incidence number in the later rather than earlier midpoint year (so in the case of 1987-1990, we plotted 1989).

After exclusions, 146 studies, from 65 countries, were included in the analysis. For 26 countries, in the analysis we have merged two or more studies with the same methodology and the same geography.

We have identified 10 groups of similar countries: Northern Western Europe, Australia/New Zealand, USA/Canada, Northern Eastern Europe, Southern Western Europe, Southern Eastern Europe, Middle East/North Africa, Hong Kong/Taiwan, Latin South America, and CIS Free Trade Area. There are also individual country data from 11 countries that cannot easily be combined into groups: China, Japan, Republic of Korea, Thailand, Maldives, Bangladesh, Mexico, US Virgin Islands, Sudan, Mali and Gabon.

**Table 1. Regional groupings and individual countries with Incidence over Time data**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group number | Region | Countries (n) | Countries (n of studies) | Number of studies, n (merged studies) | Year of studies: | Years covered with >+ 2 studies |
| 1 | Northern Western Europe | 14 | Austria (2), Belgium, Denmark (2), Finland (4), France (3), Germany (3), Iceland (2), Ireland (4), Luxembourg, Netherlands (3), Norway (3), Sweden (3)/(4 + Hagloff), Switzerland (2), UK (8) | 42 (10) | 1953-2019 | 1953-2015 |
| 2 | Northern Eastern Europe | 8 | Czechia, East Germany, Estonia (2), Hungary (2), Latvia, Lithuania (2), Poland (5), Slovakia (2) | 16 (4) | 1960-2019 | 1970-2014 |
| 3 | Southern Western Europe | 6 | Cyprus, Greece, Israel (2), Italy (5), Malta (2), Spain (5) | 16 (2) | 1984-2019 | 1985-2015 |
| 4 | Southern Eastern Europe | 9 | Armenia, Bosnia & Herzegovina, Bulgaria, Croatia, Montenegro, Nth Macedonia (2), Romania, Serbia (2), Slovenia | 11 (1) | 1974-2016 | 1982-2015 |
| 5 | Middle East/North Africa | 5 | Algeria, Jordan, Kuwait, Libya (2), Qatar | 6 (0) | 1975-2016 | 1982-2015 |
| 6 | CIS Free Trade Area | 4 | Belarus (2), Russia (4), Ukraine, Uzbekistan | 8 (0) | 1980-2014 | 1980-2011 |
| 7 | Hong Kong, Taiwan | 2 | Hong Kong (2), Taiwan (2) | 4 (1) | 1984-2017 | 1992-2014 |
| 8 | Australia/NZ | 2 | Australia (3), New Zealand (2) | 5 (1) | 1970-2019 | 1977- 2016 |
| 9 | Canada/USA | 2 | Canada (5), USA (6) | 11 (0) | 1948-2016 | 1967-2015 |
| 10 | Latin South America | 2 | Brazil (2), Chile (3) | 5 (2) | 1986-2015 | 1990- 2014 |
| 11 | China | 1 | China (6) | 6 (1) | 1980-2013 | *Use 1980-2013 as single country* |
| 12 | Japan | 1 | Japan (3) | 3 (1) | 1975-2010 | *Use 1975-2010 as single country* |
| 13 | Rep. Korea | 1 | Rep. Korea (4) | 4 (2) | 1987-2016 | *Use 1987-2016 as single country* |
| 14 | Thailand | 1 | Thailand (2) | 2(1) | 1991 -2005 | *Use 1991 -2005 as single country* |
| 15 | Maldives | 1 | Maldives (1) | 1(0) | 2009-2018 | *Use 2009-2018 as single country* |
| 16 | Bangladesh | 1 | Bangladesh (1) | 1(0) | 2011-2018 | *Use 2009-2018 as single country* |
| 17 | Mexico | 1 | Mexico (1) | 1(0) | 2000-2018 | *Use 2000-2018 as single country* |
| 18 | US Virgin Islands | 1 | US Virgin Islands (1) | 1(0) | 2001-2010 | *Use 2001-2020 as single country* |
| 19 | Sudan | 1 | Sudan (1) | 1(0) | 1991-1995 | *Use 1991-1995 as single country* |
| 20 | Mali\* | 1 | Mali\* (1) | 1(0) | 2007-2016 | *Use 2007-2016 as single country* |
| 21 | Gabon\* | 1 | Gabon\* (1) | 1(0) | 2011-2017 | *Use 2011-2017 as single country* |

\* Mali and Gabon incidence rises very likely to be inflated by improving diagnosis (i.e. less dying undiagnosed)

There are no data for large parts of Central, Western and Southern Asia, the Pacific, the Caribbean and Central America, and sub-Saharan Africa.

**Suggested approach:**

1. Develop a line of best fit of annual incidence change for each of the 10 groups, including all studies. This line can only be calculated for a particular year when there are two or more studies in that group.

This is done using \*\*\*\*

1. Extrapolations for countries in Country Group 1-10, for years without data from 1922-2021:
2. If Country X has incidence over time data:

1. Use all that data for those years for Country X
2. If there are more than one study at any particular time for Country X, average these data for that year/period of years (as long as the weighting Sensitivity Analysis on the EURODIAB data is not significant, otherwise use a weighted average when available and an unweighted average when not). (If it is significant, this will be a big spanner in the works and will slow us down).
3. When data for a particular country is not annual, interpolate the intermediate years with a straight line
4. For any period that Country X does not have incidence data, use the line of best fit for that particular Country Group to determine incidence for the missing period(s) going backwards/forwards in time.
5. When the line of best fit is not available for any period, estimated incidence remains stable at the rate it was calculated to be when the line of best fit ended.
6. If Country X does not have incidence over time data, but has known incidence at a particular time(s), project incidence before and after this/these time(s) using B. i) d) and B. 1) e) above
7. If Country X does not have any incidence data, extrapolate the data from the country assigned by the IDF Atlas method (Ogle et al. 2021)

Examples:

1. Belgium is in Group 1, and has data from 1989-2013. The group 1 line of best fit runs from 1953-2015. Therefore, this line is used to project incidence 1953-1988 and 2014-2015. Incidence remains at 1953 level before 1953 and at 2015 level after 2015.
2. Oman is in Group 5. The IDF Atlas (Ogle et al. DRCP 2021) uses an incidence from data from 1994. This figure is used for 1994, with the line of best fit data for Group 7 for 1982-2015 used to project incidence for all other years 1982-2015. Incidence remains at 1982 levels before 1982 and at 2015 level after 2015.
3. United Arab Emirates is also in Group 5. There are no incidence data. Using the extrapolation method in the IDF Atlas (Ogle et al. DRCP 2021), the incidence from Oman is used.
4. Extrapolations for all other countries for years without data from 1922-2021:
5. For individual countries with incidence over time data (Groups 11-21 in Table 1)
6. Use all available data for that country
7. there are more than one studies at any particular time for Country Y, average these data for that period (as long as the weighting Sensitivity Analysis in not significant, otherwise use a weighted average when available and an unweighted average when not)
8. For all other countries outside Groups 1-10:
9. If Country Y does not have incidence over time data but has any other incidence data (e.g. single year data), use that information for that/those year(s)
10. If the country has no incidence data, use the extrapolated figure from the IDF Atlas (Ogle et al. DRCP 2021).

**What do we do for these countries before and afterwards?**

* 1. Keep incidence at the last known rate(s) for all past and future years

or

* 1. Use a global average line of best fit. The problems with this are:
* As there are no detailed data, we do not know whether the global line of best fit is reasonable.
* In Groups 1-10, the duration of the line of best fit will vary, depending on the length of time there were at least two concurrent studies in that Group. However, if the line of best fit for all countries is used, this line will have the maximum duration (1953-2015) and lead to extrapolations for longer periods of time than nearly all other Groups.
* **My proposal is to keep incidence constant at the known value.** A number of the extrapolated studies are not recent in any case. Also, many of these countries will have high mortality rates so there will be less impact on prevalence.