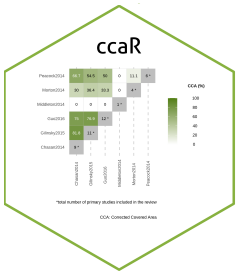
The ccaR package

Version 1.0.0

Theodoros Diakonidis Konstantinos I. Bougioukas

28/02/2022

Contents



1 Introduction

Overviews of reviews have recently become a popular approach of evidence synthesis where the unit of synthesis is the review study [1,2]. A key challenge for overview authors is to investigate the overlapping information and data across the included reviews in their overview study [3–9].

`ccaR` package provides functions for assessing and depicting primary study overlap across multiple reviews. The user needs to create an overall citation matrix in a form of a data frame. The first row of the matrix must contain the names of reviews, and the first column must contain the names of the unique primary studies included in the reviews. For each primary study, 1 should be used to indicate the review in which it has been cited otherwise it must take 0 (Table 1).

The proposed measure for assessing the overall degree of overlap in an overview of reviews is the corrected covered area (CCA) formula [3]:

$$CCA = \frac{N-r}{r \cdot c - r}$$

where N is the total number of included primary studies (including multiple occurrences of the same study) across the reviews (this is the sum of ones in the citation matrix [Table 1]); r is the number of rows (number of first occurrence primary studies [also called index publications]); and c is the number of columns (number of reviews) .

Table 1: Example of a citation matrix

	Review A	**Review B**	**Review C**	**Review D**	**Review E**
Primary study 1	1	1	1	0	0
Primary study 2	1	0	1	0	0
Primary study 3	1	0	0	1	0
Primary study 4	0	0	1	1	1
Primary study 5	0	1	1	0	1
Primary study 6	1	1	0	1	0
Primary study 7	1	0	1	1	0
Primary study 8	0	1	0	1	1

A CCA(%) value between 0% and 5% can be considered as a slight overlap, 6% to 10% as moderate overlap, 11% to 15% as high overlap, whereas values greater than 15% can be considered as a very high overlap [3].

However, a single overall CCA value for the overview of reviews may not be truly representative of the degree of overlap in some occasions [3,10]. Hennessy et al. [10] and Kho et al. [11] have raised the issue that overlaps should be investigated at the outcome level (by creating outcome matrices).

Notice that if we set c = 2 (number of reviews), the CCA formula becomes:

$$CCA_{paired} = \frac{N-r}{2 \cdot r - r} = \frac{N-r}{r} = \frac{N}{r} - 1$$

The $\text{CCA}_{\text{paired}}$ formula can be applied for each possible pair of reviews from the citation matrix (by creating a series of smaller matrices). The number of all pairs of k reviews is calculated by $k \cdot (k-1)/2$. The $\text{CCA}_{\text{paired}}$ may be used as a measure of overlap of primary studies between pairs of reviews in sophisticated graphical displays such as heatmaps [8,12].

In conclusion, the degree of overlap across reviews should be explored for the entire study and for specific outcomes of interest. The functions from `ccaR` package may be useful for methodologists and overview authors in exploring and communicating the degree of overlap in overview of reviews.

2 Installation

You can download-install the package from github and load the library:

```
#devtools::install_github('thdiakon/ccaR')
library(ccaR)
```

3 A case study

3.1 Computation of the CCA

This is a simple example of implementation of the code for the ccaR package.

First, we load the data (`test.csv`) of the working example by Miyazaki et al. study [13] . This overview summarizes the relevant evidence from six systematic reviews (SRs) [14–19] of the benefits of non pharmacological interventions for preventing type 2 diabetes mellitus (T2DM) in women diagnosed with gestational diabetes mellitus (GDM). The 6 SRs include a total of 14 unique randomized controlled trials (RCTs) relevant to the research question and eligibility criteria of the overview (Table 2).

```
DATASET <- read.csv(system.file('extdata','test.csv', package = 'ccaR'), sep = ";")
#View(DATASET)
```

Let's have a look at the data. We observe that the first row of the Table 2 contains the names of reviews (first author and year of publication), and the first column contains the names of unique primary studies (first author and year of publication) included in reviews. For each primary study, 1 is used to indicate the review in which it has been cited otherwise it takes 0. This is the default format that should have the input data for the functions `cca()` and `heat_cca()` in order to work properly.

Table 2: Citation matrix

Primary_study	Guo2016	Gilinsky2015	Morton2014	Peacock2014	Middleton2014	Chasan2014
Cheung2011	1	1	0	1	0	1
Clark2009	0	0	0	0	1	0
Ferrara2011	1	1	0	1	0	1
Hu2012	1	1	0	0	0	1
Ji2011	1	0	0	0	0	0
Kim2012	1	1	0	1	0	1
Mclntyre2012	1	1	0	1	0	1
Peterson1995	0	1	0	0	0	0
Ratner2008	1	1	1	1	0	1
Reinhardt2012	1	1	0	1	0	1
Shek2014	1	1	1	0	0	0
Shyam2013	1	1	1	0	0	1
Wein1999	1	1	1	0	0	1
Yu2012	1	0	0	0	0	0

Now we are ready to create a table with the pairwise CCA (CCA_{paired}) for each possible pair of reviews from the citation matrix using the `cca()` function:

```
cca_table <- cca(DATASET)
cca_table
#View(cca_table)
```

Table 3: CCA Table

reviews	overlap_counts	N	r	c	CCA_Proportion	CCA_Percentage
Chasan2014 vs. Gilinsky2015	9	20	11	2	0.818	81.8
Chasan2014 vs. Guo2016	9	21	12	2	0.750	75.0
Chasan2014 vs. Middleton2014	0	10	10	2	0.000	0.0
Chasan2014 vs. Morton2014	3	13	10	2	0.300	30.0
Chasan2014 vs. Peacock2014	6	15	9	2	0.667	66.7
Gilinsky2015 vs. Guo2016	10	23	13	2	0.769	76.9
Gilinsky2015 vs. Middleton2014	0	12	12	2	0.000	0.0
Gilinsky2015 vs. Morton2014	4	15	11	2	0.364	36.4
Gilinsky2015 vs. Peacock2014	6	17	11	2	0.545	54.5
Guo2016 vs. Middleton2014	0	13	13	2	0.000	0.0
Guo2016 vs. Morton2014	4	16	12	2	0.333	33.3
Guo2016 vs. Peacock2014	6	18	12	2	0.500	50.0
Middleton2014 vs. Morton2014	0	5	5	2	0.000	0.0
Middleton2014 vs. Peacock2014	0	7	7	2	0.000	0.0
Morton2014 vs. Peacock2014	1	10	9	2	0.111	11.1
Overall		43	14	6	0.414	41.4

The Table 3 includes the following columns:

1. **reviews**: pairs of reviews which are compared for overlaps of primary studies e.g., name of review X vs. name of review Y.
2. **overlap_counts**: the number of primary studies that are common between pairs of reviews.
3. **N**: the number of included primary studies (including multiple occurrences of the same study) across the pairs of reviews and overall (this is the sum of the 1s).
4. **r**: the number of rows (number of unique primary studies) for pairs of reviews and overall.
5. **c**: the number of columns (number of reviews). In the case of pairs of reviews, it equals to two.
6. **CCA_Proportion**: the proportion corrected covered area (CCA) for pairs of reviews and overall.
7. **CCA_Percentage**: the percentage (%) corrected covered area (CCA) for pairs of reviews and overall, rounded in one decimal digit.

In this working example the overall CCA(%) is equal to 41.4% which is indicated in the last row in Table 3.

3.2 Visualization of the pairwise CCA(%) with a heatmap

3.2.1 CCA(%) heatmap where reviews are ordered alphabetically

We can visualize the degree of overlap of primary studies between pairs of reviews (Figure 1).

```
heat_cca(DATASET)
```

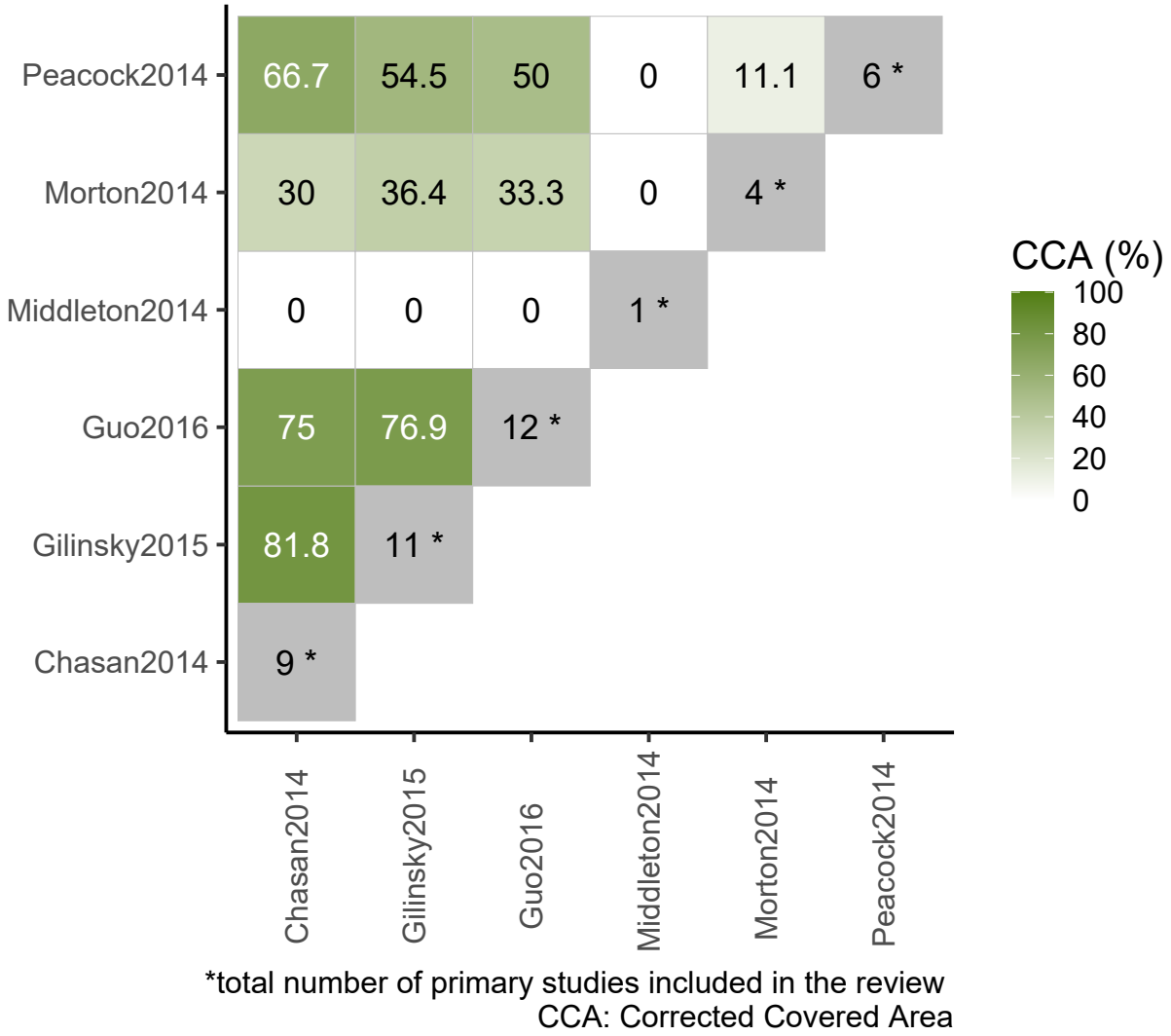


Figure 1: A simple example of a %CCA heatmap

We observe that the reviews have been ordered alphabetically. The color-coded tiles within the triangular matrix demonstrate the CCA(%) for the pairs of reviews. The diagonal, gray-colored cells indicate the total number of primary studies included in each review (Figure 1).

We can also customize the above ggplot graph with the `fontsize` argument that controls the font size of the numbers in the tiles and the `chroma` argument that allows us to change the color of the heatmap. Furthermore, we can modify theme elements if we load the `ggplot2` package (Figure 2).

```
library(ggplot2)

heat_cca(DATASET, fontsize = 6 , chroma = "blue") +
  theme(
    plot.caption = element_text(size = 14, margin=margin(30,0,0,0)),
    legend.title = element_text(size = 14, face = "bold", vjust=4),
    legend.text = element_text(size = 14),
    legend.key.size = unit(1.2, "cm"),
    legend.title.align = 1.0,
    legend.text.align = 1.0,
    axis.text.x = element_text(size = 14),
    axis.text.y = element_text(size = 14),
    axis.ticks = element_blank(),
    axis.line = element_blank(),
    panel.border = element_blank(),
    panel.grid.major.x = element_line(colour = "grey80", linetype = "dashed", size = 1.0)
  )
```

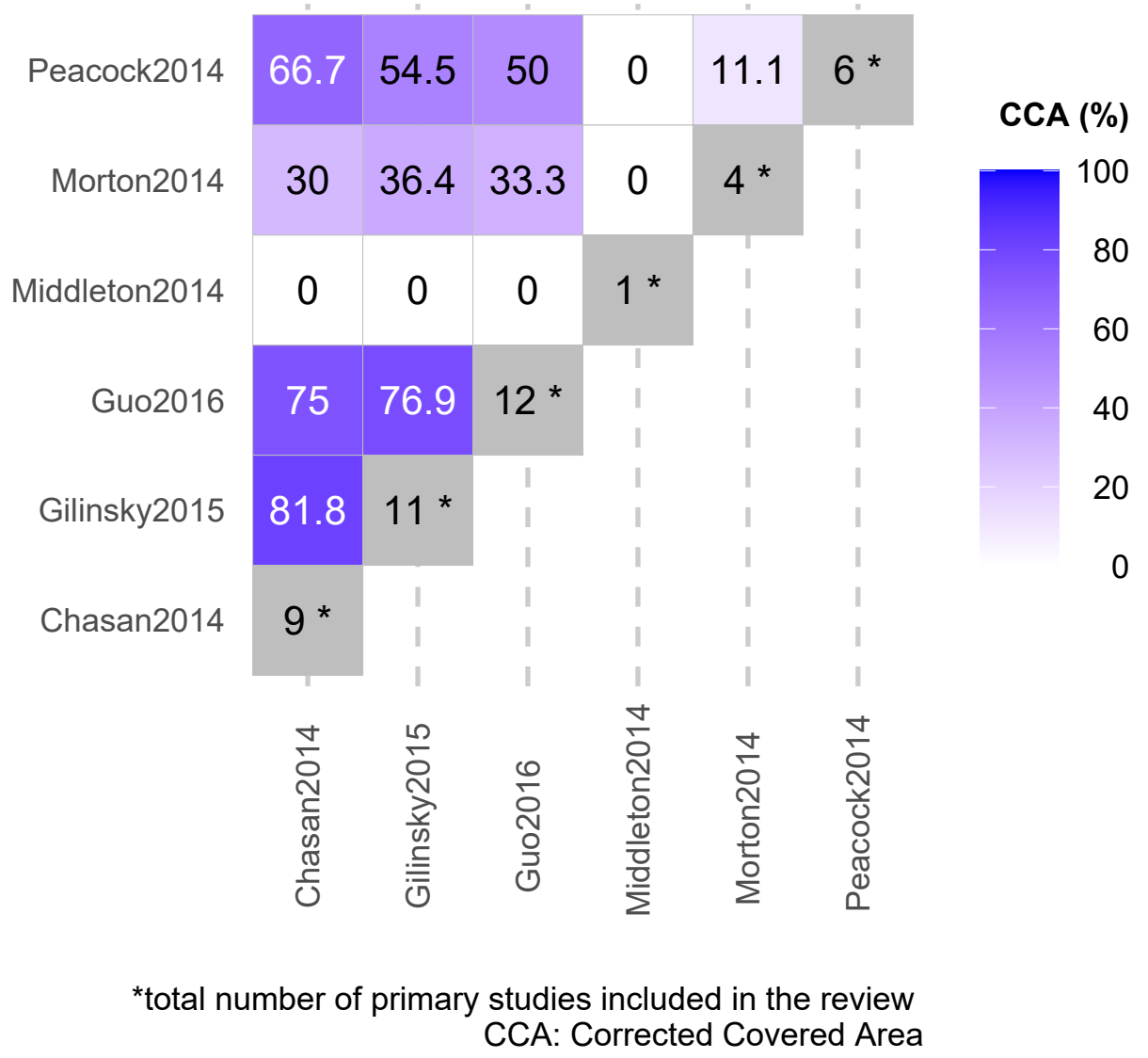


Figure 2: Customization of the %CCA heatmap

3.2.2 CCA(%) heatmap where reviews are ordered by date

If we provide a data set in which the year of publication is appeared before the name of the author of review (e.g., 2016Guo, 2014Chasan, 2015Gilinsky) then the reviews are ordered by ascending date in the heatmap, as in the following example:

```
DATASET2 <- readr::read_delim(system.file('extdata','test2.csv', package = 'ccaR'))
```

```
#> Warning in gzfile(file, mode): cannot open
#> compressed file 'C:/Users/kboug/AppData/
#> Local/Temp/RtmpKKbVlX\file4ca8316d3190',
#> probable reason 'No such file or directory'
```

```
#View(DATASET)
```

```
heat_cca(DATASET2)
```

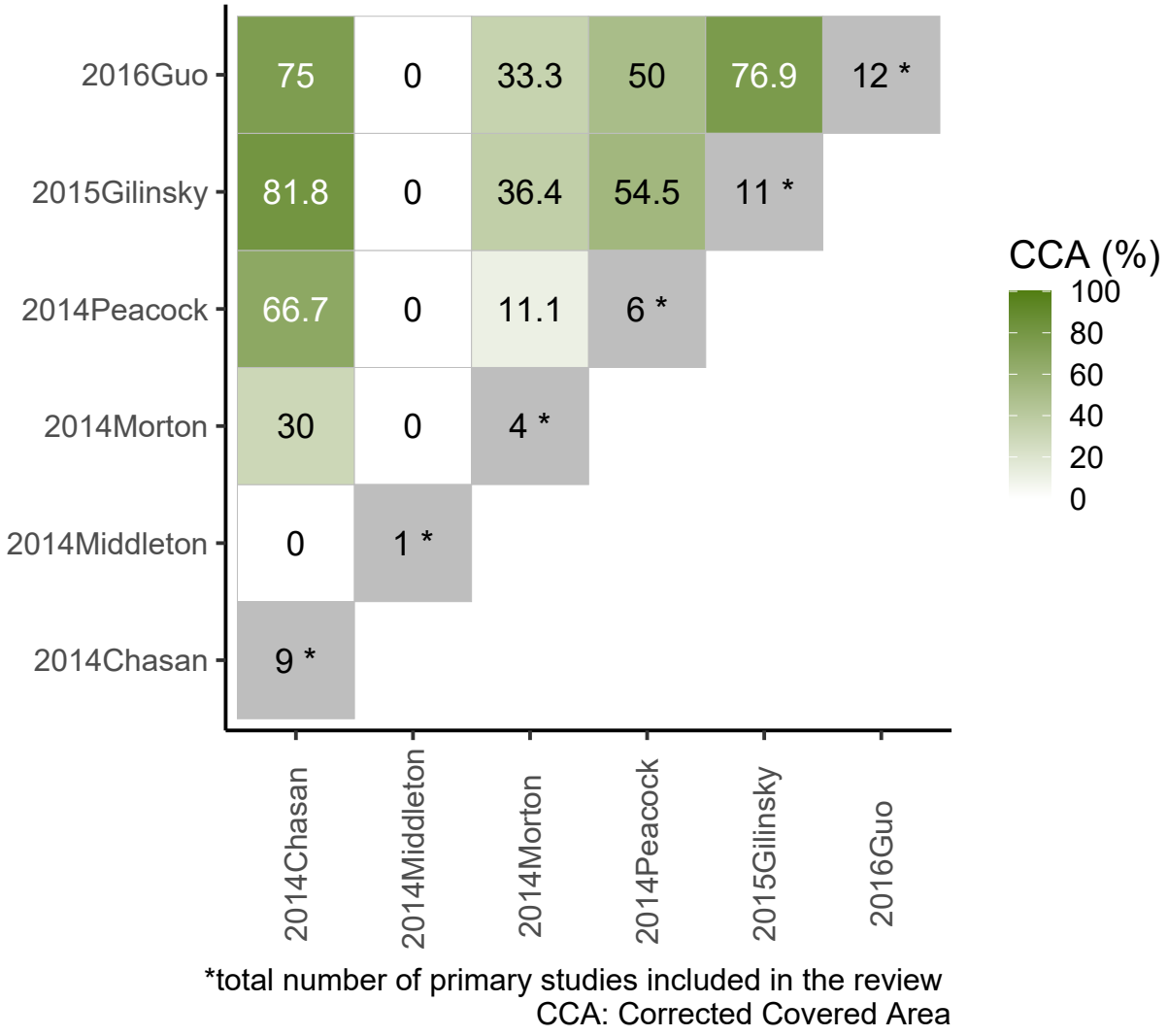


Figure 3: A simple example of a %CCA heatmap with reviews ordered by date

Note that in this case we used the `read_delim()` function from the package `{readr}` to read the `test2.csv` file.

References

1 Hartling L, Vandermeer B, Fernandes RM. Systematic reviews, overviews of reviews and comparative effectiveness reviews: A discussion of approaches to knowledge synthesis. *Evidence-Based Child Health* 2014;**9**:486–94. doi:https://doi.org/10.1002/ebch.1968

2 Bougioukas KI, Vounzoulaki E, Mantsiou CD, *et al.* Global mapping of overviews of systematic reviews in healthcare published between 2000 and 2020: a bibliometric analysis. *Journal of Clinical Epidemiology* 2021;58–72. doi:https://doi.org/10.1016/j.jclinepi.2021.03.019

3 Pieper D, Antoine SL, Mathes T, *et al.* Systematic review finds overlapping reviews were not mentioned in every other overview. *Journal of Clinical Epidemiology* 2014;**67**:368–75. doi:https://doi.org/10.1016/j.jclinepi.2013.11.007

4 Bougioukas KI, Liakos A, Tsapas A, *et al.* Preferred reporting items for overviews of systematic reviews including harms checklist: a pilot tool to be used for balanced reporting of benefits and harms. *Journal of clinical epidemiology* 2018;**93**:9–24. doi:https://doi.org/10.1016/j.jclinepi.2017.10.002

5 Bougioukas KI, Bouras E, Apostolidou-Kiouti F, *et al.* Reporting guidelines on how to write a complete and transparent abstract for overviews of systematic reviews of health care interventions. *Journal of Clinical Epidemiology* 2019;**106**:70–9. doi:https://doi.org/10.1016/j.jclinepi.2018.10.005

6 Lunny C, Brennan SE, Reid J, *et al.* Overviews of reviews incompletely report methods for handling overlapping, discordant and problematic data. *Journal of clinical epidemiology* Published Online First: 2019. doi:https://doi.org/10.1016/j.jclinepi.2019.09.025

7 Hennessy EA, Johnson BT, Keenan C. Best Practice Guidelines and Essential Methodological Steps to Conduct Rigorous and Systematic Meta-Reviews. *Applied psychology Health and well-being* 2019;**11**:353–81. doi:https://doi.org/10.1111/aphw.12169

8 Bougioukas KI, Vounzoulaki E, Mantsiou CD, *et al.* Methods for depicting overlap in overviews of systematic reviews: An introduction to static tabular and graphical displays. *Journal of Clinical Epidemiology* 2021;**132**:34–45. doi:https://doi.org/10.1016/j.jclinepi.2020.12.004

9 Pollock M, Fernandes RM, Becker LA, *et al.* Chapter V: Overviews of Reviews. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, ed. *Cochrane handbook for systematic reviews of interventions version 6.1 (updated february 2021)*. Cochrane 2021. https://training.cochrane.org/handbook/current/chapter-v

10 Hennessy EA, Johnson BT. Examining overlap of included studies in meta-reviews: Guidance for using the corrected covered area index. *Research Synthesis Methods* 2020;**11**:134–45. doi:https://doi.org/10.1002/jrsm.1390

11 Kho ME, Poitras VJ, Janssen I, *et al.* Development and application of an outcome-centric approach for conducting overviews of reviews. *Applied Physiology, Nutrition, and Metabolism* 2020;**45**:S151–64. doi:https://doi.org/10.1139/apnm-2020-0564

12 Pérez-Bracchiglione J, Niño de Guzmán E, Roqué Figuls M, *et al.* Graphical representation of overlap degree of primary studies in systematic reviews included in overviews. In: *Abstracts of the 26th cochrane colloquium, santiago, chile*. Cochrane Database of Systematic Reviews (1 Suppl 1) 2020. 151–2. doi:https://doi.org/10.1002/14651858.CD201901

13 Miyazaki C, Tanase-Nakao K, Arata N, *et al.* Nonpharmacological interventions to prevent type 2 diabetes in women diagnosed with gestational diabetes mellitus: a systematic overview. *Diabetology International* 2017;**8**:160–80. doi:https://doi.org/10.1007/s13340-017-0316-0

14 Chasan-Taber L. Lifestyle interventions to reduce risk of diabetes among women with prior gestational diabetes mellitus. *Best Practice & Research Clinical Obstetrics & Gynaecology* 2014;**29**:110–22. doi:https://doi.org/10.1016/j.bpobgyn.2014.04.019

15 Gilinsky AS, Kirk AF, Hughes AR, *et al.* Lifestyle interventions for type 2 diabetes prevention in women with prior gestational diabetes: A systematic review and meta-analysis of behavioural, anthropometric and metabolic outcomes. *Preventive Medicine Reports* 2015;**2**:448–61. doi:https://doi.org/10.1016/j.pmedr.2015.05.009

16 Guo J, Chen J-L, Whittemore R, *et al.* Postpartum lifestyle interventions to prevent type 2 diabetes among women with history of gestational diabetes: A systematic review of randomized clinical trials. *Journal of Women's Health* 2016;**25**:38–49. doi:https://doi.org/10.1089/jwh.2015.5262

17 Middleton P, Crowther C. Reminder systems for women with previous gestational diabetes mellitus to increase uptake of testing for type 2 diabetes or impaired glucose tolerance. *Cochrane Database of Systematic Reviews* Published Online First: 2014. doi:https://doi.org/10.1002/14651858.CD009578.pub2

18 Morton S, Kirkwood S, Thangaratinam S. Interventions to modify the progression to type 2 diabetes mellitus in women with gestational diabetes: a systematic review of literature. *Current opinion in obstetrics & gynecology* 2014;**26**:476–86. doi:https://doi.org/10.1097/GCO.0000000000000127

19 Peacock AS, Bogossian F, McIntyre HD, *et al.* A review of interventions to prevent type 2 diabetes after gestational diabetes. *Women and Birth* 2014;**27**:e7–15. doi:https://doi.org/10.1016/j.wombi.2014.09.002