

APPENDICES

Appendix 1: Completed CONSORT checklist for the DAA trial

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	5
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	7 – 9
	2b	Specific objectives or hypotheses	9
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	11
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	14
Participants	4a	Eligibility criteria for participants	9
	4b	Settings and locations where the data were collected	11
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	10
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	12 – 13
	6b	Any changes to trial outcomes after the trial commenced, with reasons	14
Sample size	7a	How sample size was determined	11
	7b	When applicable, explanation of any interim analyses and stopping guidelines	N/A
Randomisation: Sequence generation	8a	Method used to generate the random allocation sequence	11
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	11
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	11
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned	11

Blinding	11a	participants to interventions If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	11
	11b	If relevant, description of the similarity of interventions	N/A
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	13 – 14
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	13 – 14
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	14, 26
	13b	For each group, losses and exclusions after randomisation, together with reasons	14, 26
Recruitment	14a	Dates defining the periods of recruitment and follow-up	14
	14b	Why the trial ended or was stopped	14
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	27
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	13, 14, 27
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	15 – 18
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	15 – 18
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	18
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	N/A
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	20 – 21
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	20 – 21
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	18 – 22
Other information			
Registration	23	Registration number and name of trial registry	6
Protocol	24	Where the full trial protocol can be accessed, if	9

Funding	25	available Sources of funding and other support (such as supply of drugs), role of funders	22
---------	----	---	----

Appendix 2: Assignment of 24 pairs of data abstractors to 6 sequences and to 48 articles

							Random Sequence	
	Article 1	Article 2	Article 3	Article 4	Article 5	Article 6	Articles selected from systematic review #1	
Pair 1	A	A	B	B	C	C		Sequence 1
Pair 2	B	B	C	C	A	A		Sequence 2
Pair 3	C	C	A	A	B	B		Sequence 3
	Article 7	Article 8	Article 9	Article 10	Article 11	Article 12		
Pair 4	A	A	C	C	B	B		Sequence 4
Pair 5	B	B	A	A	C	C	Sequence 5	
Pair 6	C	C	B	B	A	A	Sequence 6	
	Article 13	Article 14	Article 15	Article 16	Article 17	Article 18	Articles selected from systematic review #2	
Pair 7	A	A	B	B	C	C		Sequence 1
Pair 8	B	B	C	C	A	A		Sequence 2
Pair 9	C	C	A	A	B	B		Sequence 3
	Article 19	Article 20	Article 21	Article 22	Article 23	Article 24		
Pair 10	A	A	C	C	B	B		Sequence 4
Pair 11	B	B	A	A	C	C	Sequence 5	
Pair 12	C	C	B	B	A	A	Sequence 6	
	Article 25	Article 26	Article 27	Article 28	Article 29	Article 30	Articles selected from systematic review #3	
Pair 13	A	A	B	B	C	C		Sequence 1
Pair 14	B	B	C	C	A	A		Sequence 2
Pair 15	C	C	A	A	B	B		Sequence 3
	Article 31	Article 32	Article 33	Article 34	Article 35	Article 36		
Pair 16	A	A	C	C	B	B		Sequence 4
Pair 17	B	B	A	A	C	C	Sequence 5	
Pair 18	C	C	B	B	A	A	Sequence 6	
	Article 37	Article 38	Article 39	Article 40	Article 41	Article 42	Articles selected from systematic review #4	
Pair 19	A	A	B	B	C	C		Sequence 1
Pair 20	B	B	C	C	A	A		Sequence 2
Pair 21	C	C	A	A	B	B		Sequence 3
	Article 43	Article 44	Article 45	Article 46	Article 47	Article 48		
Pair 22	A	A	C	C	B	B		Sequence 4
Pair 23	B	B	A	A	C	C	Sequence 5	
Pair 24	C	C	B	B	A	A	Sequence 6	

A, B & C denote three different approaches for data abstraction; see section 2.1.

Random sequence is the permuted arrangement of three approaches for data abstraction. For example, sequence 1 indicates data abstractors will collect data from 6 unique articles using AABBC approaches respectively.

Appendix 3: Additional details about methods and results for both outcomes – error proportions and time

Methods: All errors were ascertained by a computer program (Stata® Version 14, College Station, Texas) that compared the selected/entered value of a given data item to the answer key value for that data item. We defined an error as any discrepancy or difference between an entry for a data item and the answer key value for that data item. To enable these comparisons, we manually mapped abstractor-defined data items with pre-defined data items, for example, “standard deviation” (abstractor-defined) was re-coded “SD” (pre-defined). If participants abstracted more data items than were in the answer key, the additional data items were discarded and not considered as errors. We also manually double-checked each data item that had an error proportion $\geq 50\%$ in case its corresponding answer key value needed correction.

The total time taken to complete abstraction for a given article was defined as the sum of the time taken in minutes for initial abstraction(s) plus subsequent verification/adjudication. We asked each data abstractor to record the time spent on each step of data abstraction for each article: initial abstraction, verification, and adjudication (self-recorded time). These data were recorded using the online survey tool Qualtrics®. The study data abstraction system (SRDR) also automatically recorded time (automated time).

To assess the accuracy of our assessment of time, we compared self-recorded time and automated time, and noted that the self-recorded time consistently underestimated automated time, primarily because the automated clock continued to count time during breaks if the form was not saved. Our primary analysis of time focuses on the automated time.

Results for self-recorded time (see following tables): Mean times for data abstraction during the DAA trial, as captured by self-recorded time, were similar among the two verification approaches (90 minutes [range 39-229] for DAA verification; 90 minutes [range 30-285] for regular verification); times were longer with independent abstraction (142 minutes [range 59-

256] for independent abstraction) (Table A). Regardless of abstraction approach, approximately 60% of the time was spent on initial abstraction and approximately 40% on adjudication/verification. Table B provides these data aggregated across approaches.

Appendix 3-Table A: Self-recorded time (in minutes) by data abstraction approach, step of data abstraction, and review topic

	DAA verification				Regular verification				Independent abstraction			
	Step of data abstraction			Total	Step of data abstraction			Total	Step of data abstraction			Total
	Initial abstraction	Verification	Adjudication		Initial abstraction	Verification	Adjudication		Initial abstraction	Verification	Adjudication	
Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	
Topic 1	51 (36-80)	24 (11-38)	5 (0-15)	80 (61-105)	44 (20-97)	23 (5-41)	8 (0-31)	75 (39-145)	82 (40-132)	3 (0-10)	47 (22-87)	132 (65-229)
Topic 2	70 (20-210)	30 (19-60)	4 (0-30)	103 (48-229)	57 (20-140)	26 (2-50)	9 (0-30)	92 (35-172)	90 (38-182)	0 (0-0)	48 (28-90)	138 (76-255)
Topic 3	61 (20-149)	31 (10-75)	4 (0-21)	96 (45-224)	70 (18-210)	22 (10-41)	20 (0-80)	112 (30-285)	95 (69-145)	4 (0-42)	62 (24-145)	161 (98-227)
Topic 4	45 (19-72)	31 (18-65)	5 (0-20)	81 (39-145)	46 (18-113)	27 (10-55)	8 (0-28)	81 (43-136)	64 (32-153)	4 (0-20)	66 (24-190)	135 (59-256)
All Topics	56 (19-210)	29 (10-75)	5 (0-30)	90 (39-229)	54 (18-210)	25 (2-55)	12 (0-80)	90 (30-285)	83 (32-182)	3 (0-42)	56 (22-190)	142 (59-256)

Topic 1: Multi-factorial interventions to prevent falls in older adults¹⁵

Topic 2: Proprotein convertase subtilisin/kexin type 9 (PCSK-9) antibodies for adults with hypercholesterolemia¹⁶

Topic 3: Interventions to promote physical activity= in cancer survivors¹⁷

Topic 4: Omega-3 fatty acids for adults with depression¹⁸

Appendix 3-Table B: Self-recorded time (in minutes) across all approaches, by step of data abstraction and review topic

	ALL APPROACHES				
	Step of data abstraction				Total
	Initial abstraction	Verification	Adjudication	Total	
Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	Mean (Range)	
Topic 1	59 (20-132)	17 (0-41)	20 (0-87)	96 (39-229)	
Topic 2	72 (20-210)	19 (0-60)	20 (0-90)	111 (35-255)	
Topic 3	75 (18-210)	19 (0-75)	29 (0-145)	123 (30-285)	
Topic 4	52 (18-153)	21 (0-65)	27 (0-190)	99 (39-256)	
All Topics	64 (18-210)	19 (0-75)	24 (0-190)	107 (30-285)	

Topic 1: Multi-factorial interventions to prevent falls in older adults¹⁵

Topic 2: Proprotein convertase subtilisin/kexin type 9 (PCSK-9) antibodies for adults with hypercholesterolemia¹⁶

Topic 3: Interventions to promote physical activity= in cancer survivors¹⁷

Topic 4: Omega-3 fatty acids for adults with depression¹⁸

Appendix 3-Table C: Between-approach comparisons of self-recorded time across all topics

Tab	DAA Verification – Independent Abstraction			Regular Verification – Independent Abstraction			DAA Verification – Regular Verification		
	Adj. MD	95% CI	p-value	Adj. MD	95% CI	p-value	Adj. MD	95% CI	p-value
ALL Tabs	-51.8	(-66.6, -37.1)	<0.0001	-51.7	(-66.5, -36.9)	<0.0001	-0.1	(-14.9, 14.7)	0.99

Adj. MD = adjusted mean difference (adjusted for sequence and systematic review topic). The model adjusted for sequence, systematic review topic, and indicators for the approach used on the first and last article abstracted by each pair.

Appendix 4. Exploring the impact of errors on meta-analysis

PURPOSE

When data abstractors abstract incorrect data for an estimate of effect, or fail to abstract a value at all, meta-analytic summaries (i.e., point estimates and standard errors [SEs]) that are calculated might differ from meta-analytic summaries based on the “true values”. For this exploratory analysis, the “true values” were defined by an answer key (see Methods section for details). The purpose of this analysis was to explore the potential impact of these errors on meta-analyses by comparing all combinations of data abstracted by DAA trial participants to those in the answer key.

METHODS

We selected any outcome for which five or more studies reported estimates of effect for any meta-analysis that could be constructed from any of the four systematic reviews we examined. We could only construct two eligible meta-analyses, one for a continuous outcome using the mean difference (MD) as the effect measure, and one for a binary outcome using the relative risk (RR) as the effect measure. For each outcome, we computed the distribution of all possible meta-analytic summaries (MDs, RRs, and their SEs) using a random-effects model fit by the restricted maximum likelihood (ReML).^{19,20}

By the design of the DAA trial, each study was abstracted by three pairs of abstractors (Appendix 2); potential meta-analyses could thus be constructed by choosing data from one of the abstraction pairs for each study. The data abstraction chosen for a given study could be correct, incorrect, or blank because the pair failed to extract data (i.e., omission errors). With K studies, the total number of combinations would then be 3^K , but, because of the omissions, some combinations resulted in fewer than K studies with data.

Note: Because studies may have reported arm-level results, between-arm results, or both, and because we instructed abstractors to abstract both arm-level and between-arm results for each outcome of interest, we chose data for each meta-analysis using two methods. Method 1 used arm-level results if available and, if not, between-arm results. Method 2 used between-arm results if available and, if not,

derived them from the arm-level results.

We calculated the mean squared error (MSE) of the difference between each meta-analytic summary and the “true” answer key value across all the combinations for each treatment effect’s point estimate (i.e., MD or RR) and its SE. The MSE was calculated for each parameter (i.e., point estimate or SE).

Letting

- θ be the true value of the parameter (from the meta-analysis based on the answer key);
- y_{ik} be the estimate of the parameter from the i^{th} combination of results from a meta-analysis using K studies;
- \bar{y}_k be the average of the estimates of the parameter across all combinations using k studies; and
- \bar{y} be the average of the estimates of the parameter across all combinations across *all numbers of studies*.

the MSE is,

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{K_i} (y_{ik} - \theta)^2 = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{K_i} (y_{ik} - \bar{y}_k)^2 + \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{K_i} (\bar{y}_k - \bar{y})^2 + \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{K_i} (\bar{y} - \theta)^2$$

Where the three components of MSE on the right-hand side of the equation are:

- $\frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{K_i} (y_{ik} - \bar{y}_k)^2$ is the variance within a set of combinations using K studies;
- $\frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{K_i} (\bar{y}_k - \bar{y})^2$ is the variance between the means of the combinations using a fixed number of studies; and
- $(\bar{y} - \theta)$ is the bias of the overall estimate.

Based on the equation above, the MSE has three components: (i) variance within each set of combinations having the same number of studies ($y_{ik} - \bar{y}_k$), (ii) variance between the averages of sets of combinations with different numbers of studies ($\bar{y}_k - \bar{y}$), and (iii) the square of the bias, the difference between the overall mean and the “true” mean (bias) ($\bar{y} - \theta$). We calculated the percentage bias as $[(\bar{y} - \theta)/\theta] * 100$.

RESULTS

Continuous outcome (absolute change from baseline to 12 weeks in LDL-c level): Eight of the twelve studies from systematic review topic 2 (proprotein convertase subtilisin/kexin type 9 (PCSK-9) antibodies for adults with hypercholesterolemia¹⁶) provided sufficient data for a meta-analysis for this outcome. With 3 data abstractors for each study and 8 studies, there were $3^8=6,561$ combinations of data that could be used for this meta-analysis. Because data abstractors sometimes omitted an outcome (i.e., failed to abstract any data for an outcome) or did not abstract sufficient data for a meta-analysis (e.g., abstracted mean without measures of precision), many combinations resulted in fewer than 8 studies (the mean was 5.67 studies per meta-analysis [range 3 to 8]). The distribution of number of studies included in each meta-analysis is shown below:

# of studies	Freq.	Percent	Cum.
3	108	1.65	1.65
4	756	11.52	13.17
5	1,971	30.04	43.21
6	2,322	35.39	78.60
7	1,188	18.11	96.71
8	216	3.29	100.00
Total	6,561	100.00	

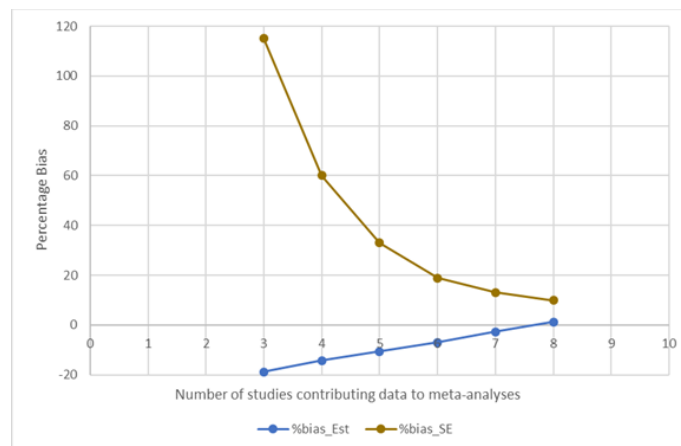
When using data from the answer key, the mean of the MDs under Method 1 was -2.09 mmol/L (95% CI: -2.64 to -1.53). The SE of this MD was 0.28. This compares to a mean of -1.92 mmol/L averaging across all 6,561 combinations of abstractions. The largest and smallest possible meta-analytical estimates by selecting the most extreme results were -2.27 mmol/L (95% CI: -3.06 to -1.48) and -1.63 mmol/L (95% CI: -2.45 to -0.81), respectively.

The overall MSE for the MDs derived from the iterative meta-analysis was 0.0562 (combining within-group variance of 0.0206, between-group variance of 0.00728, and squared bias of 0.0283). The bias was thus 0.1682. Therefore, **the percentage bias in the continuous outcome resampling meta-analysis estimate was -8.0%** (i.e., $[0.1682/-2.09]*100$).

The overall MSE for the SE derived from the iterative meta-analysis was 0.0125, which included a within-group variance of 0.00367, between-group variance of 0.00259, and squared bias of 0.00620. The bias was thus 0.07874 and so **the percentage bias in the continuous outcome resampling meta-analysis standard error was +28.1%** (i.e., $[0.07874/0.28]*100$).

Figure A below plots the percentage bias in the continuous outcome meta-analysis estimates and SEs for iterations based on different realized values of K (i.e., different numbers of studies). As seen in Figure A, for the continuous outcome, as the number of studies included in each meta-analysis increases, the percentage bias in the estimate and the percentage bias in the SE both decrease.

Appendix 4 Figure A: Percentage bias in meta-analysis estimates and standard errors for iterations based on different realized values of K (i.e., different numbers of studies) for the continuous outcome



The results for analysis Method 2 were similar to those for analysis Method 1 (results not presented).

Binary outcome (participants with at least one fall by 12 months): Ten of the twelve studies in systematic review topic 1 (multi-factorial interventions to prevent falls in older adults¹⁵)

provided sufficient data for a meta-analysis for this outcome. With 3 data abstractors for each study and 10 studies, there were $3^{10}=59,049$ combinations of data that could be used for this meta-analysis. There was a mean of 6.67 studies per meta-analysis (range 4 to 9) for both analysis Methods 1 and 2. The distribution of number of studies included in each meta-analysis is shown below:

# of studies	Freq.	Percent	Cum.
4	972	1.65	1.65
5	6,804	11.52	13.17
6	17,739	30.04	43.21
7	20,898	35.39	78.60
8	10,692	18.11	96.71
9	1,944	3.29	100.00
Total	59,049	100.00	

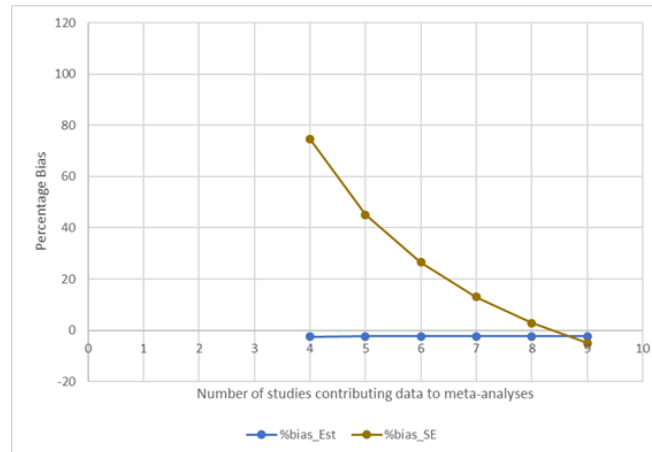
When using data from the answer key, the mean of the RRs (averaged on the log scale and then exponentiated) using analysis Method 1 was 0.93 (95% CI: 0.84 to 1.03). This compares to a mean of 0.91 averaging across all 59,049 combinations of abstractions. The largest and smallest possible meta-analytical estimates by selecting the most extreme results were 0.86 (95% CI: 0.74 to 0.99) and 0.97 (95% CI: 0.87 to 1.09), respectively.

The overall MSE for the RRs derived from the iterative meta-analysis was 0.000921, a sum of the within-group variance of 0.000477, between-group variance of 0.000000099, and squared bias of 0.000445. The bias was thus -0.02109 and **the percentage bias in the binary outcome resampling meta-analysis estimate was -2.3%** (i.e., $[-0.02109/0.93]*100$).

The overall MSE for the SEs derived from the iterative meta-analysis error was 0.000247, comprising a within-group variance of 0.0000975, between-group variance of 0.0000563, and squared bias of 0.0000934. The bias was thus 0.00966 and **the percentage bias in the binary outcome resampling meta-analysis standard error was +19.3%** (i.e., $[0.00966/0.05]*100$).

As seen in Figure B, for the binary outcome, as the number of studies on which meta-analysis iteration is based increases, the percentage bias decreases.

Appendix 4 Figure B: Percentage bias in resampling meta-analysis estimates and standard errors for iterations based on different realized values of K (i.e., different numbers of studies) for binary outcome



The results for analysis Method 2 (not shown) were similar to those for analysis Method 1.

LIMITATIONS

This analysis is based on the two outcomes that fulfilled our eligibility criteria (i.e., five or more studies reporting results eligible for meta-analysis of the outcome).

CONCLUSIONS

For both the outcomes we analyzed, i.e., one continuous and one binary outcome, any errors in the data that were abstracted during the DAA trial would not make a sizable impact on meta-analytic summaries (i.e., estimates and standard errors). For the continuous outcome, the percentage biases in the estimate and standard error of the mean difference (MD) were -8.0% and 21.8%, respectively. Similarly, for the binary outcome, the percentage biases in the estimate and standard error of the relative risk (RR) were -2.3% and 19.3%, respectively. Omission of data during abstraction was the largest contributor to errors; meta-analytic

summaries based on fewer studies were more biased than summaries based on larger numbers of studies.