Assembling a raw data meta-analytic database : the joys and sorrows

Sue Duval, PhD and David Jacobs, PhD
on behalf of the CODA Study Group
June 24, 2004
DMICC, Bethesda

Overall scientific aims

- Individual studies attempt to make inferences setting up experimental contrasts that pertain to the hypothesis. Nevertheless, observed findings are subject to random variation that could lead the inference astray.
  - Rule out chance using standard statistical tests and p-values
  - Difficult to test consistency of findings across a variety of settings in a single study
- Hence a goal of meta-analysis is to enhance inference by increasing power and by assessing consistency of findings across studies

CODA aims

- To use individual participant data meta-analysis [MIPD] methods to address the following outstanding issues in diabetes epidemiology:
  - What simple anthropometric indices most closely predict the risk of T2DM in adults?
  - Do ethnicity and other factors modify that prediction?
  - Is the association of these anthropometric indices with cardiovascular disease morbidity and mortality exacerbated by their association with T2DM?
  - Is it possible to predict several diabetes-related risk states [IFG and IGT (collectively referred to as IGM), undiagnosed diabetes, and diabetes incidence] using noninvasive or minimally invasive methods?
  - Should the screening tools differ by ethnicity?

Types of meta-analysis/Terminology

- Systematic Review
  - Exhaustive exploration, critical evaluation and synthesis of all the unbiased evidence
- Meta-analysis of the published literature
  - Sometimes called MAL, MPL
  - Exhaustive exploration, critical evaluation and quantitative synthesis of all the unbiased evidence from published reports
  - Combination of the published results of a number of studies
- Meta-analysis of individual participant data [retrospective]
  - Sometimes called MIPD, IPD, MAP, 'pooled' analysis in epidemiology
- Meta-analysis of individual patient data [prospective]

What is an MIPD?

- Requires GOOD international collaboration
- Involves the central collection, checking and analysis of individual participant data
- Includes all studies, published and unpublished
- Has been described as the "yardstick" and "gold standard" of systematic reviews

Why MIPD?

- Analyses based on published data can give different answers to an MIPD due to:
  - Exclusion of studies
  - Exclusion of participants
  - Time-point of analysis
  - Length of follow-up
  - Method of analysis
  - Inadequate reporting
  - Covariate adjustment

MIPD vs MAL

- Advantages
  - Data
    - More information
    - Inclusion of extended databases from published studies
    - Inclusion of data from unpublished studies
  - Better standardization of information
    - Categorization of eligible participants
    - Outcomes, exposures, covariates
    - Definition of follow-up period and censoring criteria
  - Analysis
    - Better time-to-event analyses
    - Better adjusted/multivariate models
    - Evaluation of subgroup effects
  - Interpretation
    - Assessment of heterogeneity
    - Assessment of sampling bias in specific studies
  - Other
    - Establishing international networks of collaborating investigators

MIPD vs MAL

- Disadvantages
  - Data
    - May not be available from all published studies
  - Interpretation
    - Potential conflicts with collaborators regarding findings
  - Resources
    - Substantial effort and infrastructure required to
      - Develop and administer protocol
      - Collect, manage and analyse data
      - Communicate with collaborators

The CODA Project MIPD

- Joys
  - Great detail about broader questions can be answered eg waist vs BMI in fine tuned demographic groups, and in different ranges of adiposity
- Sorrows
  - Very detailed questions cannot be answered this way eg utility of thigh circumference since only a few studies have measured it

MIPD as a study design

- The biggest advantage is personal relationships with the collaborators
  - Accomplished by phone, email, meetings. The face-to-face meeting is the most powerful
  - Life becomes easier with a positive track record of valuable publications
  - Potential for encouraging increased collaborative thinking
- Other benefits to a collaborative effort
  - More complete identification of studies
  - More balanced interpretation of the results
  - Wider endorsement and dissemination of results
  - Better clarification of further research
  - Collaboration on further research

Key principles

- All data sent to the data management site are
  - held securely and treated in the strictest confidence
  - not used in any publication without the permission of the responsible collaborator
- All published reports of the meta-analysis results are and will be
  - in the name of the CODA Study Group
  - circulated to all members of the group for comment and approval before publication

Running an MIPD

- Ultimate aim is to obtain accurate, up to date data for all individuals in all relevant studies
- Most effort is required to establish and maintain collaboration, and to process data
- Care must be taken when merging different datasets
  - Protocols similar enough?
  - Source populations poolable?
- Least problematic area might be the analysis itself, although there are challenges

Resource requirements

- Time — Several years
- Expertise — Clinical, Scientific, Statistical, Data Management, Computing, Administrative
- Money — ~ $1,000 per study; Travel for meetings
- Staff — ~ 80% of budget

Organisational structure

- Central site [University of Minnesota] comprises local staff and relevant experts
- Most decisions taken by local staff
- A larger Steering Group may be set up to advise on strategic issues
- All participating studies should be members of the collaborative group
- Writing committee comprised of interested parties

Formal protocol

- Formal protocol is vital
- Allows a meta-analysis to be designed with the same rigor as any other study
  - specify rationale behind project
  - set out main aims and objectives
  - specify a priori hypotheses and methods
- Useful in clarifying issues, identifying potential problems and explaining the project to collaborators
- Publication [most vital to collaborators]
  - Must ensure that individual studies have first rights on publication of their data and that all studies' local review rules are followed

Identifying Studies

- Inclusion criteria
  - Baseline measures of age, sex, race/ethnicity and one or more anthropometric indicators of obesity such as WC, BMI, or waist to hip ratio.
  - Follow-up for T2DM incidence. Cross-sectional studies with newly diagnosed cases of impaired glucose tolerance or T2DM are also included.
- Utmost importance to identify and include as many relevant studies as possible
- If missing trials are numerous or unrepresentative they could affect the meta-analysis results in an important way
- WHO MONICA, DECODE, DECODA, Medline searches, screening of abstracts of major international diabetes conferences, personal communication with experts in the field

Include published and unpublished studies

- Considerable evidence that 'positive' studies are more likely to be published than 'negative' studies [publication bias]
- Publication of an apparently sound manuscript does not guarantee the quality of the data
- IPD allows the meta-analysis team to perform more extensive 'peer review'
- Can avoid a number of reporting biases such as publication bias, duplicate reporting bias, outcome reporting bias, participant exclusion bias and follow-up bias

Establishing collaboration

- Initial letter inviting collaboration, but not yet asking for data
  - main aims and objectives
  - importance of the collaborative group
  - publication policy
  - collaborative group policy
  - confidentiality of data
- Specific questions relating to study eligibility
  - Short questionnaire used
- Asked for study protocol

Contacting collaborators - practical problems

- Older studies
  - investigators moved/retired
  - cooperative groups disbanded
  - data lost
- Contact 2nd, 3rd, 4th,.....authors
- Contact national institutions and agencies
- Geographical problems
- Some disinterested, too busy

CODA 'ideal' variables

- Class 1: Variables that can be measured using questionnaire/self-report only
  - Age, sex, race, ethnicity, family history of diabetes, gestational diabetes, gestational hypertension, hypertension and use of prescription anti-hypertensive medications, ethanol consumption, education level, smoking, menopausal status, hormone replacement, statin use, coronary heart disease status
- Class 2: Clinical variables that do NOT require drawing blood
  - Weight, height, waist circumference, hip circumference, other obesity measurements, blood pressure (systolic and diastolic)
- Class 3: Clinical variables that require drawing blood, but do NOT require a provocative challenge OGTT
  - Fasting or non-fasting total cholesterol and HDL cholesterol, fasting triglycerides, glucose, insulin
- Class 4: Clinical variables that require OGTT
  - 2 hour glucose level, 2 hour insulin level
- Dependent variables
  - IGM (excluding diabetes), previously undiagnosed diabetes, previously diagnosed diabetes, diabetes incidence

Data Collection & Transfer

- Flexibility of format
  - Suggested coding provided
  - Accept whatever the collaborator can send
  - Data managers can reformat data
- Assistance
  - Supply data forms
  - Financial
- Flexibility of data transfer methods
  - FTP to secure site, email, CD

Maintaining Contact with Collaborators

- Ideally
  - Regular correspondence
  - Meetings
  - Manuscripts

Data Checking

- Read trial protocol and check that it is consistent with eligibility criteria
- Seek the most recent follow up possible
- Check received data is correct
  - Not to centrally police studies
  - Check for missing data
  - Compare data received with publications
  - Perform range checks and flag outliers to be verified
  - Check consistency across variables within a participant
  - Tabulate data and send simple summary statistics to collaborator for verification
- Ask questions!
- Trim datasets to allow pooling [eg restrict age/trim outliers]

Quality Scoring

- MIPD usually have a simple binary score
  - study is included
  - study is excluded
- Quality scoring systems largely relate to randomized trial publications
- MIPD allows for very detailed checking
- Aim is to 'clean' all data sets to be of high quality

Rejecting a Study

- Discuss issues in detail with collaborators
- Most problems are due to error
- If study has to be excluded, it should be mentioned briefly in the MIPD publication (depending on the exclusion reason, it is desirable to present sensitivity analyses including the questionable or problematic studies)

MIPD analyses

- Individual participant data used
- Analyses stratified by study
- IPD does not mean that all individuals are combined into a single mega-study [sometimes also called pooling]
- One approach is to re-analyse each study and combine summary estimates using traditional MAL

'Survival' Analysis

- MAL
  - Restricted to analysis at a fixed point in time, or to a series of fixed time-points
- MIPD
  - Uses individual survival times to calculate expected number of events
  - Takes account of censoring
  - Useful when time-to-event is important
  - Produces survival curves

Subgroup Analyses

- MIPD may achieve sufficient power to allow the assessment of whether associations are larger or smaller in any subgroup.
- Should be a reasonable biological explanation for any observed interactions
- Usual cautions apply
- Can aid interpretation of the results
- Pre-specify, interpret cautiously
- Look for consistency across studies

If IPD are not available

- Aggregate unpublished data
- Aggregate published data
  - Weighting?
- Wait
- Leave study out and rely on what you have

Collaborators' Meeting?

- Is a meeting of collaborators necessary?
  - Email, phone-calls, conferences for subsets
- Together with group publication makes the project collaborative
- Gives the collaborators the first opportunity
  - to discuss the results
  - to challenge the analyses
  - to discuss the interpretation and implication of the results
- Sets a deadline to which everyone involved has to work towards
- Incentive to collaborate
- Role of meeting
  - To present the results
  - To discuss the methods, results and implications
  - To discuss publication
  - To decide what to do next
    - Further analysis
    - Additional projects

Publishing Results

- MIPD are collaborative projects
- Carried out on behalf of a collaborative group
  - Collaborators
  - Writing Group
- Published on behalf of the group
  - AOCTG (BMJ 1991)
  - SMAC (Lancet 1997)
  - EBCTCG (Lancet 1996; 1998a,b; 2000)
  - ABC (Lancet 2003)

Collaborators

- The joys
  - A subset is very interested, want to be on writing committees.
- The sorrows
  - A bigger subset is apathetic, too busy. They don't usually respond.
  - Rarely encounter collaborators who actually want hands-on or writing

Our experience so far

- The joys

  "Many thanks for sending the beautiful and interesting results and for the poster and slide presentations. I have found it is a very good representation of the analysis. Especially you have been working very hard on this, so thank you again and congratulations. Moreover I completely rely on you for the future manuscript."

  "Nice work, nice answers to good questions. Congratulation."

  " …, well done! That part of the story seems fairly solid."

- The sorrows?
  - 17/29 email responses to our ADA abstract [although most within 36 hours]
  - 9/33 email responses to our ADA poster [although several collaborators attended the poster session]

Sharing data …. ethics

- Reluctance to share data fairly universal and understandable
- Collaborators have no real interest [personal or institutional gain] in sharing data
- $1000 for their work to provide data and documentation – one study declined on the basis of too little monetary reward
- DETECT-2

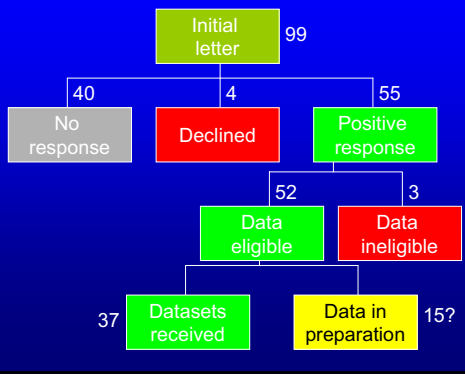
Interaction with collaborators

- Each interaction is bulky
  - 50 collaborating studies, multiple people in charge, some studies have multiple layers of review
  - Multiple iterations to check data, clarify questions
  - Getting their attention can be difficult
- Expansion of the database difficult
  - Update [add a new exam or follow-up]
  - Expand [add new variables]

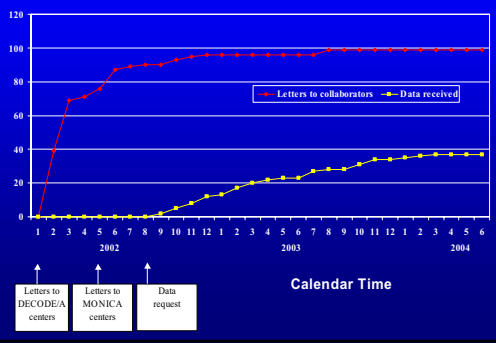
Our overall experience

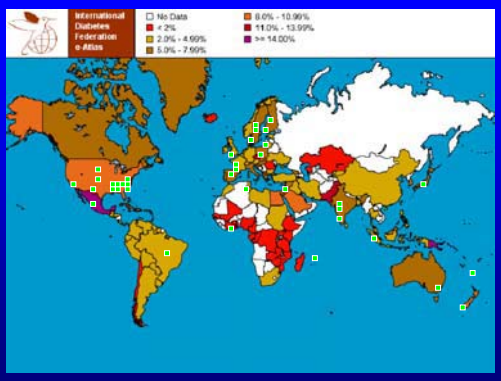
- Very positive in at least TWO ways
  - Quick response to output intended for abstracts
  - Glowing remarks concerning our ADA poster
- Not so interested in data management questions
  - Our collaborators are not data managers, so this is natural

Accumulation of the CODA database



Accumulation of the CODA database





PROSPECTIVE STUDIES

Study	Baseline years	Country	N	% women	Age range	Mean FU [years]	WC-BMI Corr	BMI Mean	WC Mean	% with newly dx DM	DM rate per 10,000
ARIC	1986-1990	USA	15,792	55	44-66	8	0.87	27.2	95.6	4.6	160.2
CARDIA	1985-1986	USA	5,115	54	17-35	13	0.86	25.7	80.6	0.4	82.7
ELY	1990-1992	UK	1,040	57	40-67	9	0.71	25.6	83.0		62.4
FIN-MON	1987,1992	Finland	11,997	53	25-64	9	0.78	26.5	86.7	1.7	16.2
FRAMINGHAM	1995	USA	3,197	53	22-79	4	0.82	26.5	88.5		64.2
GOTEBORG	1968-1970	Sweden	1,462	100	38-61	24	0.85	23.9	72.6		32.4
HIROSHIMA	1994-1996	Japan	907	52	30-80	4		23.5		8.1	325.5
IOWA	1986	USA	41,836	100	52-71	10	0.82	26.7	87.1		58.0
IRAS	1992-1994	USA	1,624	56	39-69	5	0.81	28.2	89.9	11.4	287.3
JACDS	1983-1988	USA	658	47	34-75	10	0.84	24.1	85.8	2.9	210.9
MAURITIUS	1987	Mauritius	5,078	53	25-75	9	0.84	23.7	76.3	6.8	223.7
MEXCITY	1990-1992	Mexico	2,282	59	29-67	6	0.86	28.0	95.9	3.2	137.4
NAS	1961-1968	USA	2,214	0	21-81	14	0.85	25.9	93.6	4.3	31.7
NAURU	1987	Nauru	868	56	19-81	7	0.86	34.5	97.0	17.7	313.0
NSWED-MON	1986-1999	Sweden	6,947	51	25-74	8	0.79	25.5	87.2	1.7	59.9
NURSES	1986-1987	USA	52,468	100	39-67	12	0.81	24.5	78.6		29.9
PARIS	1967-1972	France	7,746	0	43-53	4	0.88	25.3	91.1	2.8	120.1
RANCHO	1984-1987	USA	2,480	56	23-96	8	0.75	24.8	83.6	3.6	122.1
SAHS	1979-1988	USA	5,158	57	24-69	7	0.79	27.7	89.6	3.1	160.1
SHS	1989-1992	USA	4,549	59	44-75	7	0.89	29.7	101.6	16.1	401.7
ULSAM	1970-1973	Sweden	2,322	0	50	21	0.86	25.0	87.2	1.7	65.5

CROSS-SECTIONAL STUDIES

Study	Baseline years	Country	N	% women	Age range	WC-BMI Corr	BMI Mean	WC Mean	% with newly dx DM
AusDiab	1999-2000	Australia	11,247	55	25-95	0.80	26.9	90.6	2.3
CATALONIA	1994	Spain	2,217	56	29-91	0.69	26.1	88.7	2.7
CUPS	1996-1998	India	1,262	56	20-90	0.68	23.0	78.4	4.8
CURES	2001-2002	India	25,902	51	20-90	0.59	23.0	81.1	7.7
GHANA	1998	Ghana	577	55	25-91	0.85	25.9	85.2	2.7
KAUNAS-MON	1992-1993	Lithuania	1,239	51	35-64	0.79	27.5	87.6	1.5
LILLE-MON	1995-1996	France	1,195	50	36-67	0.81	26.3	90.0	3.4
NUDS	1999-200	India	11,215	53	20-96	0.60	23.6	81.5	5.1
OULU55	1990-1991	Finland	831	57	55		26.4		2.2
POL-MON	1992-1993	Poland	466	53	43-73	0.79	26.4	91.7	3.4
SINGAPORE	1998	Singapore	4,723	54	18-69	0.81	28.0	81.3	6.0
TUNIS	1995	Tunis	862	60	31-93	0.81	24.0	93.9	8.5
WORKNZ	1988-1990	New Zealand	6,577	28	40-78	0.78	27.2	90.6	2.9

Basic characteristics of the current database

- 37 studies
- 22 follow-up studies
- 20 countries
- ~115,000 persons with glucose measurement at baseline (over age 30)
- ~1.4 million person years for follow-up diabetes
- A few exceptions
  - Studies without information on waist circumference
  - Studies with restricted sample

Diabetes outcomes in the CODA database

(n of studies: 30 with men and 31 with women)

CODA GROUP

***Writing Group*: Sue Duval [Chair], David Jacobs, Gabriela Vazquez, Karri Silventoinen**

***Studies and Collaborators:* Australian Obesity, Diabetes, and Lifestyle Study (AusDiab): J Shaw; Atherosclerosis Risk in Communities (ARIC): J Stevens; Coronary Artery Risk Development in Young Adults (CARDIA): P Schreiner; Catalonia Study (CATALONIA): C Castell; Chennai Urban Population Study (CUPS): V Mohan; Chennai Urban Rural Epidemiology Study (CURES): V Mohan; MRC Ely Study (ELY): N Wareham; Finland-MONICA (FIN-MONICA): J Tuomilehto ; Framingham Offspring Study (FRAMINGHAM): P Wilson; Non-communicable Diseases Survey in 1998 (GHANA): A Amoah; Women in Goteborg Study (GOTEBORG): L Lissner; OGTT Follow up Study (HIROSHIMA): C Ito; IOWA Women's Health Study (IOWA): A Folsom; Insulin Resistance Atherosclerosis Study (IRAS): L Wagenknecht; Japanese American Community Diabetes Study (JACDS): E Boyko; Kaunas MONICA Study (KAUNAS-MON): S Domarkiene; MONICA Lille Study (LILLE-MON): J Dallongeville; Mauritius Non-communicable Disease Study (MAURITIUS): J Shaw; Mexico City Diabetes Study (MEXCITY): M Stern; Normative Aging Study (NAS): P Cassano; Nauru Study (NAURU): J Shaw; Northern Sweden MONICA Study (NSWED-MON): M Eliasson; National Urban Diabetes Study (NUDS): A Ramachandran; Nurses' Health Study (NURSES): D Feskanich; Oulu 55 (OULU55): S Keinanen-Kiukaanniemi; Paris Prospective Study (PARIS): M A Charles; Poland-MONICA (POL-MON): A Pajak; Rancho Bernardo Study (RANCHO): E Barrett Connor; San Antonio Heart Study (SAHS): M Stern; Strong Heart Study (SHS): B Howard; Singapore National Health Survey (SINGAPORE): J Cutter; Tunis CVD Study (TUNIS): F Harzallah; Uppsala Study of Adult Men (ULSAM): B Zethelius; Workforce Diabetes Survey (WORKNZ): R Scragg**

Acknowledgments

The CODA Project is supported by the US Centers for Disease Control and Prevention, Division of Diabetes Translation