Challenge in utilizing CDISC standards for integrating individual participants' data from cohort studies in Japan: the EPOCH-JAPAN study

Anna Tsutsui PhD.[1)], Yoshitaka Murakami PhD.[1)], and on behalf of the EPOCH-JAPAN Research Group

[1)] Department of Medical Statistics, Faculty of Medicine, Toho University

Meet the Speaker

Dr. Anna Tsutsui

Title: Assistant Professor

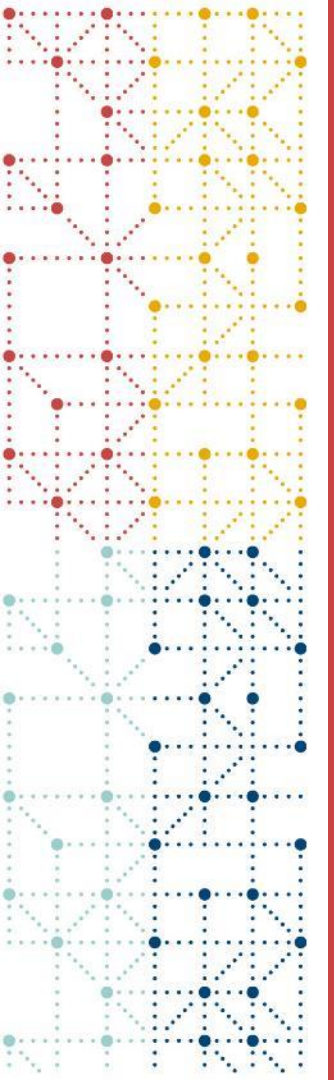
Organization: Toho University

She is an assistant professor at the Department of Medical Statistics, the Faculty of Medicine, Toho University. With a decade of experience, she specializes in implementing CDISC Standards and conducting statistical analysis. Following her retirement, she obtained a Ph.D. in Health Science. After completing a one-year post-doctoral fellowship, she assumed her current position.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the authors and do not necessarily reflect the official policy or position of CDISC.*

- *The authors have no real or apparent conflicts of interest to report.*



Agenda

1. Our project and databases
2. Our consideration and challenges in data integration

Our project and databases

Cohort study

One type of non-experimental epidemiologic study

"All subjects in a source population are classified according to their exposure status and followed over time to ascertain disease incidence" (KJ Rothman, 2021)

Example:

- Framingham Heart Study (1948-, U.S.)
- Cohorts of atomic bomb survivors in Hiroshima & Nagasaki (1950-, Japan)

EPOCH-JAPAN

Regional/Occupational Cohorts
(Three national-scale cohorts are not shown)

Large-scale meta-analysis study project:

Evidence for Cardiovascular Prevention from Observational Cohorts in Japan

Cited from epi-c.jp (2022)

- Initiated in 2005
- Integrated personal data from **240,000** individuals across **17 observational cohorts** in Japan (epi-c.jp, 2022)
  - ✓ Health Japan 21 (second term) (2013)
  - ✓ Management guidelines for high blood pressure in Japan (2014)

New project of EPOCH-JAPAN

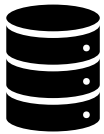
(Leader: Prof. Y Murakami (23FA1006), FY2023-25)

To develop interactive tools that
promote behavioral change in individuals and
to generate evidence to solve individual health issues based on ...

Expanded existing database (Baseline measurements)

New integrated longitudinal databases:

- Including ≧**15,000** individuals from **eight** cohorts
  - Annual health check-ups
  - High-precision measurements available
  - Capturing disease onset over time



cdisc

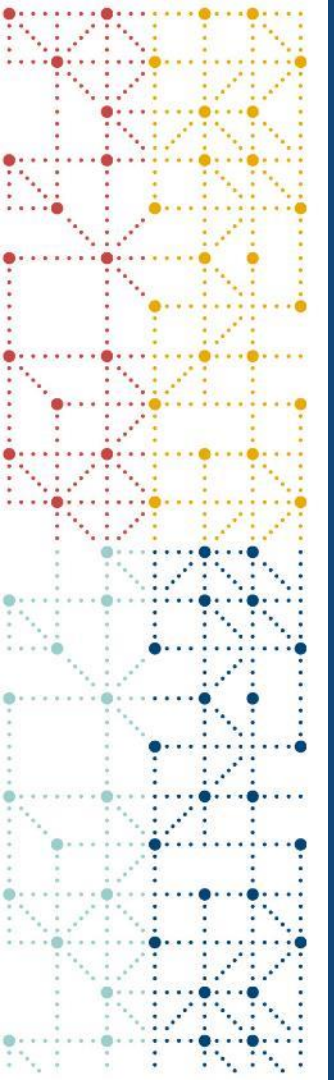
Study objective and methods

Our study objective was to develop new integrated databases for long-term use **in global data standards format.**

Methods:

- We consulted with CDISC experts.
- Subsequently, we integrated the databases by using CDISC-controlled terminology and the ADaM standards.

ADaM, Analysis Data Model



Our consideration and challenges in data integration

Cohort Data Characteristics

Male
1
0

Sex
1
2

Female
0
1

性別	Sex
男	M
女	F

In many cases (though not always)…

- Each cohort utilizes its own terminology and the data structures of cohorts are different from each other.
- Data are often composed in the local language, Japanese.
- Simple data structure
  - ✓ Minimal necessary data are collected
  - ✓ Researchers manage data
  - ✓ One record per subject per time-point
- Groups (TRTxxA), analysis value (AVAL), & covariates vary for research questions even if the same cohort data are used.

PTNO	Year	WBC	RBC	..
1	1998	xx	xx	xx
1	1999	xx	xx	xx
2	1998	xx	xx	xx
2	1999	xx	xx	xx

Our questions on data integration

In clinical trials, CDISC standards are used to integrate data by:

* If necessary

In cohort studies…

1. Are there any standards other than CDISC standards?
2. Are CDISC standards suitable?
3. Although cohort data often have a simple data structure, is the SDTM conversion necessary in advance?

SDTM, Study Data Tabulation Model

Consultations

We consult with multiple CDISC experts at …

- CDISC 2023 Japan Interchange and
- Monthly meetings in the SDTM team of CDISC Japan User Group

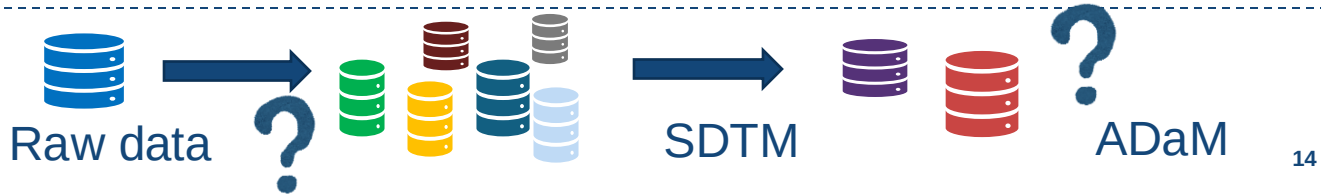
1. Are there any standards other than CDISC standards?
   No. / No comment.
2. Are CDISC standards suitable?
   Yes. **CDISC has draft guidance** for SDTM for observational studies. (next slide).
3. Although cohort data often have a simple data structure, is the SDTM conversion necessary in advance?
   Yes. Would it not be better to unify variable names or terminologies?

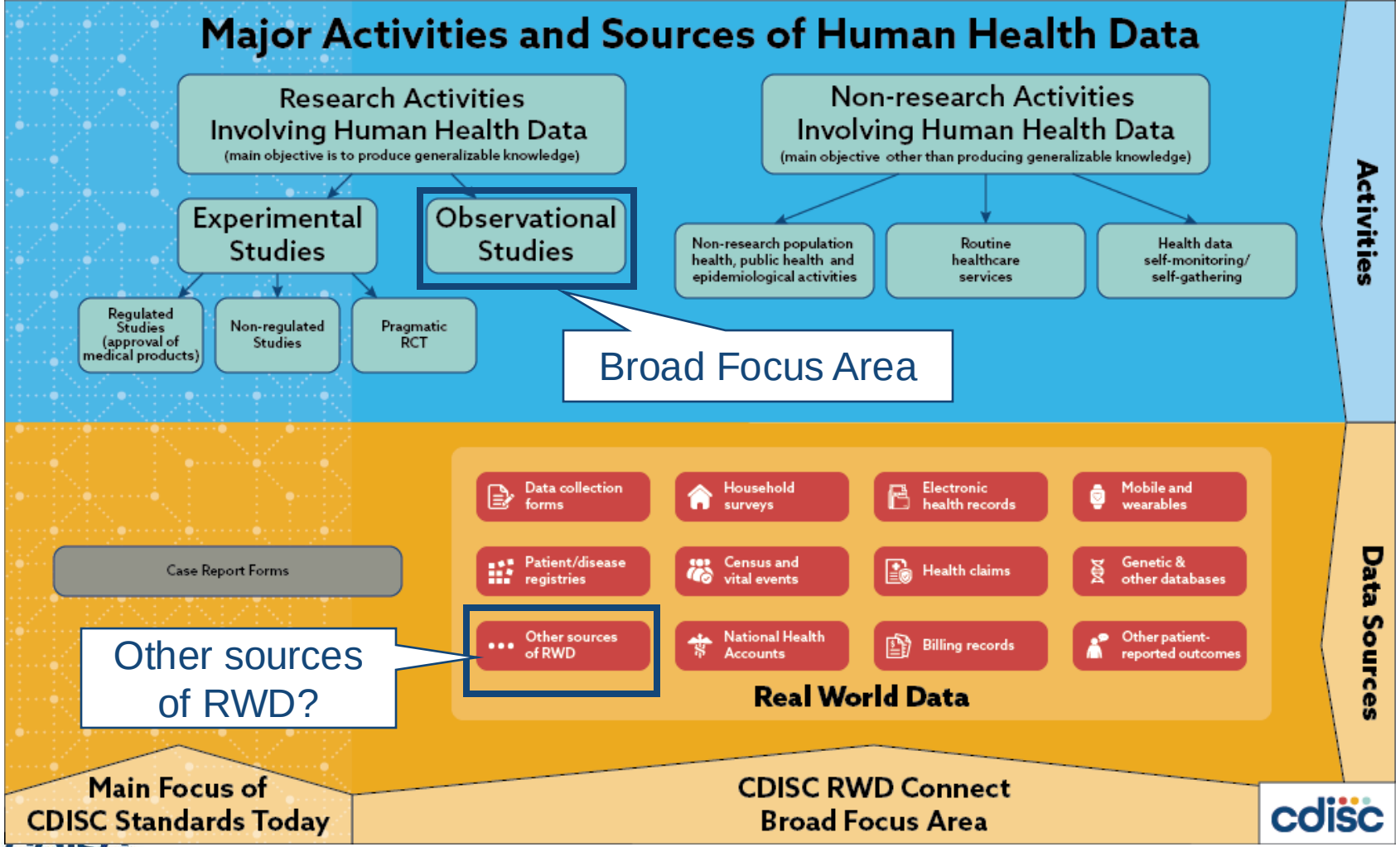
CDISC guidance for observational studies (Not a standard)

"Considerations for SDTM Implementation in Observational Studies and Real-World Data v1" (Feb 2024)

- Provides strategies for handling commonly encountered **issues** when using the SDTM for observational studies and RWD
- Scope
  - Observational Studies: **Cohort** and Case-control study
  - Real-World Data: External control arm (ECA) or comparator cohort

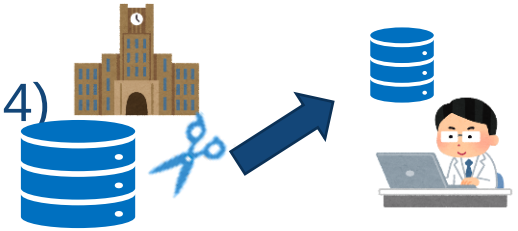
Owing to the nature of SDTM standards, splitting the data by subject is necessary, which was anticipated to be resource-intensive.





CDISC (2024) Real World Data. https://www.cdisc.org/standards/real-world-data

Actual data obtained (N=5, as of March 2024)



We obtained cohort datasets from each collaborative research institution & found that data are limited to essential data necessary for our study.

The data obtained from each cohort were classified into 3 patterns:

1. One dataset including all the information (N=3).
2. One baseline dataset and one post-baseline dataset (N=1).
3. Five datasets per year and one outcome data (N=1).

Additionally, when creating the ADaM specifications for longitudinal data, we found that applying them to BDS is rather far away from the “analysis-ready” principle (next slides).

BDS, Basic Data Structure

Original Cohort-data ("analysis-ready")

PTNO	Year	SBP	DBP	..
1	1998	150	130	xx
1	1999	xx	xx	xx
..	..	..	..	..

Total Cholesterol
Non-HDL Cholesterol
…
Concomitant Medications for xxxx
Concomitant Medications for yyyy
…

BDS data

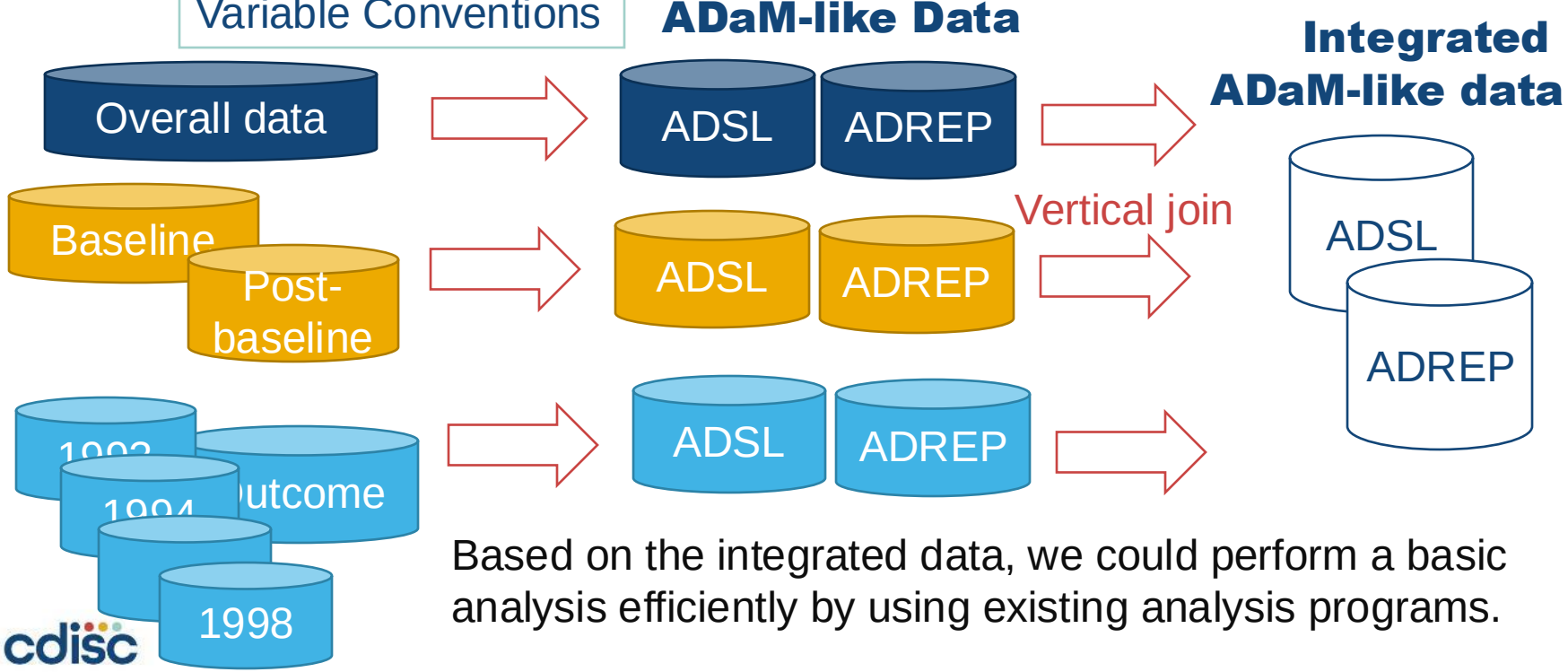
(Not analysis-ready and needs re-transpose and recombination of covariates and other variables)

SUBJID	AVISITN	PARAMCD	AVAL	..
1	1998	SYSBP	150	xx
1	1999	DIABP	130	xx
..	..	..	..	..

Data flow

ADSL: almost in conformance with standards
ADREP: BDS-like data
but one record per subject per year

Terminology,
Variable Conventions



Based on the integrated data, we could perform a basic analysis efficiently by using existing analysis programs.

Our data conformance summary

We followed the following standards to the extent possible:

- CDISC Controlled Terminology
- ADaM standards except for BDS
  - 3.1* ADaM Variable Conventions
  - 3.2* ADSL Variables

Conventions on …
- General Variable
- Timing Variable
- Flag Variable Conventions

Variable Naming Fragments

We deviated from the following CDISC standards:

- SDTM standards (Not used)
- BDS in ADaM standards
  - 3.3* ADaM Basic Data Structure (BDS) Variables

*Chapter numbers in ADaM IG version 1.3

Discussions 1

We integrated cohort databases **partially** based on CDISC standards.

The data integration and basic meta-analysis were performed more efficiently with high quality.

- CDISC-controlled terminology almost covers one used in our studies.
- Variable conventions in the ADaM standards saved time.
- However, BDS has a different structure with common cohort data, which was anticipated not to be analysis-ready for our meta-analysis.

（Matsuzaki, 2023)

A study integrated the data from three cohorts **fully** based on CDISC standards.

- Observational studies and RWD include a variety of data sources, and the applicability of these standards may differ for each data source.

Discussions 2

- 45% of CDISC member organizations hope CDISC providing more information on representing RWD (CDISC, 2024).
- Even with the demand for meta-analysis, guidance for the implementation of cohort data is limited.
- Researchers may have to consider other standards like Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

Despite the above limitations,
CDISC standards are still beneficial for data integration in cohort studies and will be desirable to develop more guidance in urgent work.

Ethical considerations

- This research plan obtained ethical approval from the Ethics Committee of the Faculty of Medicine of Toho University on July 31, 2023 (A23053), and additional approval for the addition of institutions to the plan on November 9, 2023 (A23075_A23053).

Funding Sources

- This research was supported by a grant–in–aid from the Ministry of Health, Labor and Welfare, Health and Labor Sciences research grants, Japan (Research on Health Services: H17–Kenkou–007; Comprehensive Research on Cardiovascular Disease and Life–Related Disease: H18–Junkankitou [Seishuu]–Ippan–012; H19–Junkankitou [Seishuu]–Ippan–012; H20–Junkankitou [Seishuu]–Ippan–013; H23–Junkankitou [Seishuu]–Ippan–005; H26-Junkankitou [Seisaku]-Ippan-001; H29–Junkankitou–Ippan–003; 20FA1002, and 23FA1006).

cdisc

Acknowledgements

- The Evidence for Cardiovascular Prevention from Observational Cohorts in Japan (EPOCH–JAPAN) Research Group is composed of the following investigators. Co-Chairperson: Hirotsugu Ueshima Tomonori Okamura, Yoshitaka Murakami
- Executive committee: Hirotsugu Ueshima, Yutaka Kiyohara, Toshiharu Ninomiya, Yutaka Imai, Takayoshi Ohkubo, Hiroyasu Iso, Isao Muraki, Kazumasa Yamagishi, Akiko Tamakoshi, Yoshihiro Miyamoto, Yoshihiro Kokubo, Katsuyuki Miura, Sachiko Tanaka, Akiko Harada, Shigeyuki Saitoh, Hirofumi Ohnishi, Ichiro Tsuji, Atsushi Hozawa, Hideaki Nakagawa, Masaru Sakurai, Michiko Yamada, Yoshimi Tatsukawa, Kiyomi Sakata, Kouzo Tanno, Akihiko Kitamura, Masahiko Kiyama, Yuji Shimizu, Akira Okayama, Shizukiyo Ishikawa, Hiroshi Yatsuya, Takeo Nakayama, Fujiko Irie, and Toshimi Sairenchi.

cdisc

References

- CDISC (2024) Real World Data. https://www.cdisc.org/standards/real-world-data (accessed on 28 April 2024).
- CDISC (2024) Considerations for SDTM Implementation in Observational Studies and Real-World Data v1. https://www.cdisc.org/standards/real-world-data (accessed on 28 April 2024).
- epi-c.jp (2022) EPOCH-JAPAN. http://www.epi-c.jp/e402_1_0001.html (accessed on 28 April 2024).
- KJ Rothman et, al. (2021) Modern Epidemiology 4th Edition, Lippincott Williams & Wilkins, Philadelphia.
- Matsuzaki K, Kitayama M, et, al. A Pragmatic Method to Integrate Data From Preexisting Cohort Studies Using the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model: Case Study. JMIR Med Inform. 2023 Dec 21;11:e46725. doi: 10.2196/46725.

cdisc

Thank You!

Feel free to contact me at
anna.tsutsui@med.toho-u.ac.jp