Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

“Data Matters!”

Winter Course in Clinical Research

2025

OVERVIEW OF RESEARCH DESIGN & BIOSTATISTICS

A practical introduction to data management, the role of data and statistics in clinical research – and how to find help!

Examples of gathering quality evidence for impact will be discussed and the role data and statistics plays in research design. This will be followed by an interactive session where participants' research needs will be explored.

Dr Caitriona Ryan

Associate Professor in Statistics and
Trial Design School of Medicine TCD

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

TOPICS:

- Introduction
- Generating evidence for IMPACT
- How to structure and store data on MS Excel?
- The role of statistical thinking and how to get help!
- Interactive discussion
- Experimental vs observational research

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Introduction

Introduction

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

Prof Caitriona Ryan
Associate Professor in Biostatistics

PUBLIC HEALTH AND PRIMARY CARE / STAFF / *CAITRIONA RYAN*

Caitriona Ryan

Caitríona Ryan is the Associate Professor in Statistics and Trial Design in the School of Medicine at Trinity College Dublin and the Wellcome-HRB Clinical Research Facility in St. James's hospital. Her qualifications include a BSc in mathematical science, MSc in statistics and PhD in statistical clustering of network data at UCD. She furthered her career as a Research Associate in QUT, Brisbane, Australia where she developed an adaptive experimental design algorithm for biological process models. She returned to Ireland in 2015 and has since worked as an academic statistician in Maynooth University with the SFI i-Form Centre for Advanced Manufacturing and the Trinity Centre for Ageing and Intellectual Disability.

Examples of Caitríona's work include supporting statistical planning and analysis of clinical trials, modelling multimorbidity networks and optimising the 3D printing of personalised medicines and medical implants at point of care. She is also interested in longitudinal data analysis, anomaly detection and Bayesian inference. She led the data team for the latest wave of data collection in the Intellectual Disability Supplement to the Irish Longitudinal Study on Ageing, managing the entire data lifecycle; from software testing and data wrangling to statistical analysis in collaboration with theme leads across many disciplines including Nursing and Midwifery, Dental Science and Pharmacy & Pharmaceutical Sciences. Dr Ryan is a member of the Irish Statistical Association

Statistical Expertise

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

REGULATED ROLE

ICH-GCP - any regulated CTIMP must have a qualified trial statistician involved.

Declaration of Helsinki - all medical research should be rigorous and methodologically sound and avoid research waste.

Trial statistician

Reporting

Data safety monitoring

IN-HOUSE SUPPORT

- Building expertise in biostatistics for clinical trial design
- Collaborating with PIs
- Providing study design support at grant application stage – **biostatistician now requirement of HRB ILCT**
- Improve statistics -> improve quality of research
- Not just sample size calculations!!!

RESEARCH & INNOVATION

- **Optimize evidence-based and personalized care.**
- Personalised RCTs: N-of-1 trials
- Platform trials, Emulation trials
- Bayesian adaptive trials
- Impact: Bridge between clinical practice and cutting-edge science / research.

CALM

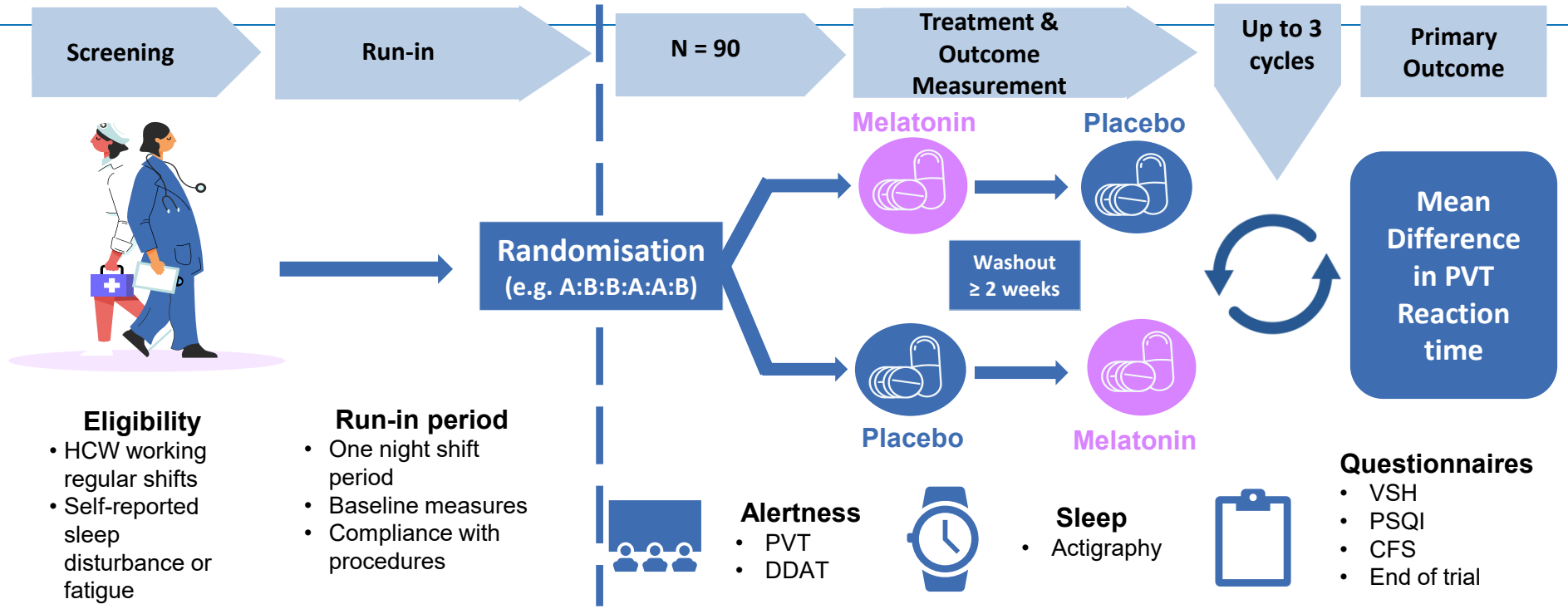
Circadian ALignment with Melatonin
A Randomised, Double-Blind, Placebo-Controlled N-of-1 Crossover Trial of Melatonin for Healthcare Staff Working Night Shifts

Dr. Caitríona Ryan, Statistician

CALM

Circadian ALignment with Melatonin

A Randomised, Double-Blind, Placebo-Controlled N-of-1 Crossover Trial of Melatonin for Healthcare Staff Working Night Shifts



Day	1	2	3	4	5	6	7	8	9
Morning	Baseline PVT, DDAT, VSH, PSQI	End Shift 1 Dose 1	End Shift 2 Dose 2, PVT, DDAT	End Shift 3 Dose 3, PVT, DDAT	End Shift 4	Recovery 1	Recovery 2	Recovery 3	End: PVT, DDAT, VSH, CFS, PSQI
Evening	Start Shift 1	Start Shift 2, PVT, DDAT	Start Shift 3, PVT, DDAT	Start Shift 4, PVT, DDAT	Dose 4, PVT, DDAT, VSH, CFS	Dose 5, PVT, DDAT	Dose 6, PVT, DDAT	Dose 7, PVT, DDAT	

CFS Chandler Fatigue Scale DDAT Double Digit Addition Test PSQI Pittsburgh Sleep Quality Scale PVT Psychomotor Vigilance Test VSH Vernon Shyder Halpern Sleep Scale

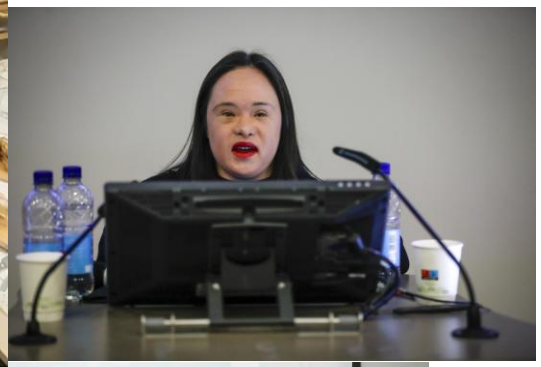
Dr. Caitríona Ryan, Statistician

Experience in School of Nursing and Midwifery

TCAID - IDS-TILDA - Statistician and Data Team Lead

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

- Software testing
- Fieldworker training
- Regular data checks throughout Interviews
- Data entry for Pre-Interview Questionnaires
- Data cleaning, merging
- Analysing the data in collaboration with theme leads for the Wave 5 Report
- Mentoring PhD students
- Supporting research output
- Generating evidence for change



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Generating Evidence for Impact

Ageing and Intellectual Disability

Importance of Investigating Ageing of People with ID

The Intellectual Disability Supplement to The Irish Longitudinal Study on Ageing (IDS-TILDA)

LINK YOUR RESEARCH TO WHO SUSTAINABILE DEVELOPMENT GOALS!

KNOW YOUR RESEARCH AREA! What are the gaps? What is your motivation? What are the issues you want to highlight?

Wealth of international research highlighting:

- Significant health inequalities compared to the general population (e.g. Cooper, Melville & Morrison, 2004)
- Higher levels of health needs than the general population (e.g. Wilson & Haire, 1990; Kapell et al, 1998) and these often go unrecognized and unmet (e.g. Lennox & Kerr, 1997)
- Data suggesting lower longevity in Ireland (46 yrs) compared to cohorts of persons with ID identified in the U.S. (66 yrs) and the UK (58-74 years) (Lavin et al, 2006; Bittles, 2002; Janicki, 1999)

- People with ID do not access health promotion and health screening services to the same extent as peers without disability.

IDS-TILDA: Research Objectives

The Intellectual Disability Supplement to The Irish Longitudinal Study on Ageing (IDS-TILDA)

To understand the health characteristics of people ageing with an intellectual disability;

BE CLEAR ABOUT YOUR RESEARCH OBJECTIVES!

To examine the service needs and health service utilisation of people ageing with an intellectual disability;

To identify disparities in the health status of adults with an intellectual disability as compared to TILDA findings for the general population;

To support evidence-informed policies, practices and evaluation.

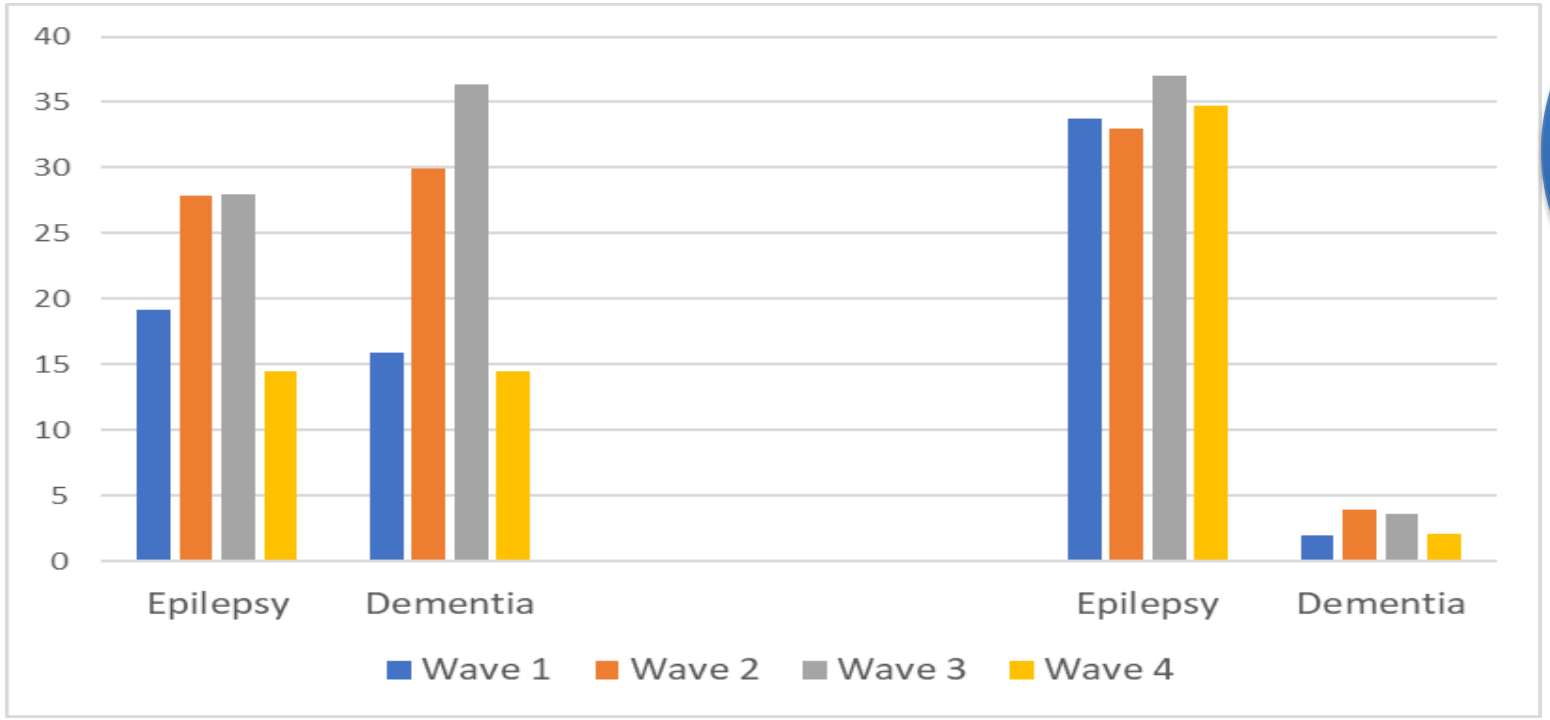
Brain Health

USE SIMPLE GRAPHS / VISUALISATIONS!

The Intellectual Disability Supplement to The Irish Longitudinal Study on Ageing (IDS-TILDA)

People with Down syndrome

People with ID from other aetiologies



54.7 mean age of onset for dementia in those with Down syndrome

65.8 mean age of onset for dementia for ID of other aetiology

Opportunity: National Intellectual Disability Memory Service

IMPACT!

The odds of a participant with dementia having Down syndrome, is 10.68 times as likely as the odds of having dementia, but other ID aetiology

Dr. Caitríona Ryan, Statistician

Bone Health

The Intellectual Disability Supplement to The Irish Longitudinal Study on Ageing (IDS-TILDA)

6%
of men had received a doctor's diagnoses for osteoporosis **Wave 2**

90%
of men had objective evidence of poor bone health **Wave 2**

USE SIMPLE DESCRIPTIVE STATISTICS! - they can tell an important story

IMPACT!

Opportunity:
Bone Health Intervention Programme
https://www.getwiseid.eu/#/home

Men in IDS-TILDA were **12 times** more likely to present with osteoporosis than men in the TILDA study

Dr. Caitríona Ryan, Statistician

PPI in IDS-TILDA

by people with an intellectual disability

TRINITY PPI IGNITE OFFICE

The Intellectual Disability Supplement to The Irish Longitudinal Study on Ageing (IDS-TILDA)

Inputting on protocols

Training fieldworkers

Disseminating results

Translating Findings

Research is everybody's business...
including people
an intellectual disability

IMPACT:
PPI engagement:
Nothing about us without us

Dr. Caitríona Ryan, Statistician

IMPACT - Research as a Driver for EVIDENCE-BASED Change!

The Intellectual Disability Supplement to The Irish Longitudinal Study on Ageing (IDS-TILDA)

- Dataset has been described as ‘a national treasure’.
- IDS-TILDA Data collection is mapped to policy changes and questions added / rested to track the effect of policy changes.
- Regular discussions / collaboration with Government Departments.
  - The National Statistics Board met with IDS-TILDA and CSO in Government buildings.
- PI Prof Mary McCarron has been invited to the Commission on Care for Older People
  - examining the health and social care services and supports for older people across the continuum of care and make recommendations for their strategic developments.

Impact beyond SJH - motivation to publish!

Contribute to research driven clinical practice

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

How to Structure and Store your Data

How to structure and store data on MS Excel? - Example of a 'BAD' Dataset

Wellcome - HRB Clinical Research Facility at St. James's Hospital

Patient ID	Name	Age	Gender	Diagnosis	Date of Admission	Blood Pressure (mmHg)	Heart Rate (bpm)	Medications Prescribed	Discharge Date	Follow-up Appointment	Doctor's Notes
1	Mary-Jane Smith	29	Female	Hypertension #$%%	20/05/2023	140/90	80	Lisinopril, Hydrochlorothiazide	20/06/2023	15/07/2023	Needs to reduce stress!!!
2	Bob Lee	thirtY	Male	Asthma**	10/06/2023	110/70	90	Albuterol, Fluticasone	15/06/2023	forgot	Asthma medication working.
3	John D	45	male	"Heart Attack, MI	12/07/2023	150/100	60	Metoprolol, Aspirin, Clopidogrel	20/07/2023	15/08/2023	Needs urgent heart surgery!!!!!!!
4	Sarah Lee	32	Female	Diabetes Type 2	01/08/2023	120/80	78	Insulin, Metformin**	15/08/2023	01/09/2023	Continue diet control, good improvement.
5	(Blank Name)	27	(Unknown)	Rheumatoid arthritis!!!	10/09/2023	130/85	85	Methotrexate, Prednisone	20/09/2023	05/10/2023	Needs pain management!
6	Emily Trenton	55	Female	Chronic Fatigue Syndrom???	03/10/2023	125/75	72	Doxycycline, Lyrica	15/10/2023	01/11/2023	Fatigue levels fluctuating.
7	Liam O'Conner	60	Male	Hyperlipidemia, High Chol.	15/10/2023	140/95	76	Statins, Fish oil	20/10/2023	05/11/2023	Improve cholesterol with diet!
8	N/A	34	Female	Migraines / Stress issues?	02/11/2023	135/80	82	Sumatriptan, Beta blockers	10/11/2023	N/A	Migraine frequency decreased.
9	Jason Black	38	Male	Sleep Apnea **Risk	01/12/2023	145/85	72	CPAP machine, Melatonin	10/12/2023	01/01/2024	Continue with CPAP treatment.
10	Lily Green	50	Female	"Osteoarthritis	05/12/2023	140/92	80	Celecoxib, Calcium supplements	15/12/2023	10/01/2024	Chronic pain management needed.

Example of a ‘BAD’ Dataset

Wellcome - HRB Clinical Research Facility at St. James’s Hospital

Patient ID	Name	Age	Gender	Diagnosis	Date of Admission	Blood Pressure (mmHg)	Heart Rate (bpm)	Medications Prescribed	Discharge Date	Follow-up Appointment	Doctor's Notes
1	Mary-Jane Smith	29	Female	Hypertension #$%%	20/05/2023	140/90	80	Lisinopril, Hydrochlorothiazide	20/06/2023	15/07/2023	Needs to reduce stress!!!
2	Bob Lee	thirtY	Male	Asthma**	10/06/2023	110/70	90	Albuterol, Fluticasone	15/06/2023	forgot	Asthma medication working.
3	John D	45	male	"Heart Attack, MI	12/07/2023	150/100	60	Metoprolol, Aspirin, Clopidogrel	20/07/2023	15/08/2023	Needs urgent heart surgery!!!!!!!
4	Sarah Lee	32	Female	Diabetes Type 2	01-08-2023	120/80	78	Insulin, Metformin**	15/08/2023	01/09/2023	Continue diet control, good improvement.
5	(Blank Name)	27	(Unknown)	Rheumatoid arthritis!!!	10/09/2023	130/85	85	Methotrexate, Prednisone	20/09/2023	05/10/2023	Needs pain management!
6	Emily Trenton	55	Female	Chronic Fatigue Syndrom???	03/10/2023	125/75	72	Doxycycline, Lyrica	15/10/2023	01/11/2023	Fatigue levels fluctuating.
7	Liam O'Conner	60	Male	Hyperlipidemia, High Chol.	15/10/2023	140/95	76	Statins, Fish oil	20/10/2023	05/11/2023	Improve cholesterol with diet!
8	N/A	34	Female	Migraines / Stress issues?	02/11/2023	135/80	82	Sumatriptan, Beta blockers	10/11/2023	N/A	Migraine frequency decreased.
9	Jason Black	38	Male	Sleep Apnea **Risk	01/12/2023	145/85	72	CPAP machine, Melatonin	10/12/2023	01/01/2024	Continue with CPAP treatment.
10	Lily Green	50	Female	"Osteoarthritis	05/12/2023	140/92	80	Celecoxib, Calcium supplements	15/12/2023	10/01/2024	Chronic pain management needed.

- Inconsistent Data Entry
- Unstructured and Non-Standardized Text
- Inconsistent Date Formats
- Missing or Incomplete Data
- Unorganized Doctor's Notes
- Free text with random capitalization, exclamation marks, and incomplete thoughts
- Inconsistent Medication Entries

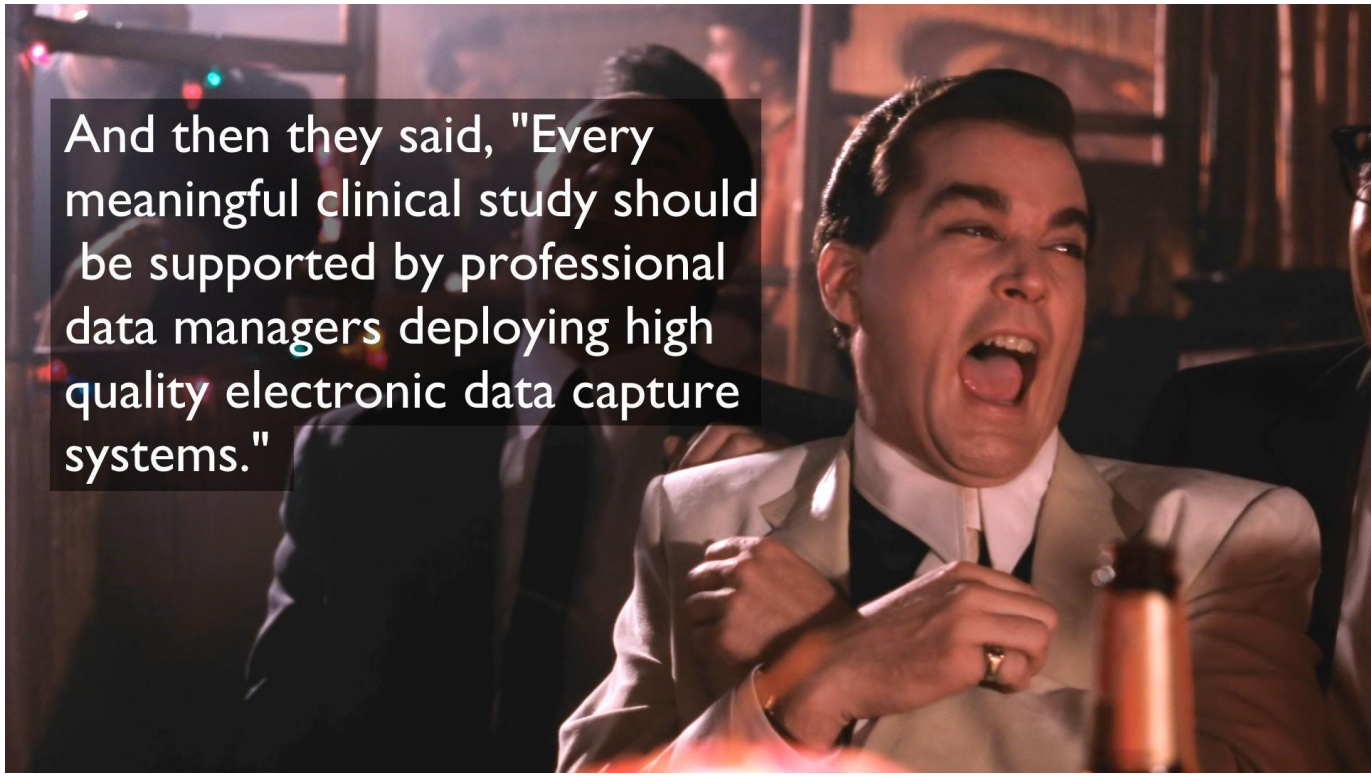
How to Structure and Store Data on MS Excel?

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

Blogpost from Darren Dahly of CRF Cork: Simple tips for recording data in spreadsheets

AIM to collect and manage data in a manner that **avoids data errors**.

We want to ensure that our datasets can **facilitate easy analysis and reporting** of results.



To meet these objectives, just follow these simple rules:

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

Easy Rules to Structure and Store Data on Excel

1. Rule 1. Use rows and columns carefully – each row is single patient. Each column is a characteristic of the patients. No blank columns. No blank rows.

2. Rule 2. Columns should include only one type of information e.g. words or numbers or dates. Each entry in the column should have the same format and unit.

3. Rule 3. The first row of your spreadsheet should contain short informative column names.

Patient_ID	Gender	Date	Age	Level	Heart_rate
1	Male	24.01.25	13	High	75
2	Female	26.01.25	29	Low	80
3	Female	01.02.25	11	Low	85
4	Female	02.02.25	41	Medium	90

What About Missing Observations?

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

RULE: Just leave it blank!

Don't use 9,99,999 etc .

It's a good idea to add a second column to explain why it's missing.

Patient_ID	Gender	Date	Age	Level	Heart_rate	Heart_rate_miss
1	Male	24.01.25	13	High	75	
2	Female	26.01.25	29	Low	80	
3	Female	01.02.25	11	Low		machine out of service
4	Female	02.02.25	41	Medium	90	

Data Dictionary

- It's a good idea to create a second dataset that is *about* your dataset.

- The same rules from above apply, but now each row will reflect one of your variables, and each column will be some information about those variables.

Variable	Unit	Label
ID		Patient id
Sex		Sex
Date		Day measurements taken
Age	years	Age
Level		Level of adherence
Heart_rate	bpm	Heart rate
Heart_rate_miss		Reason for missingingess
BMI	kg/m2	Body Mass index

Other Considerations

Humans make mistakes. We want to minimize those. Many data collection tools you might use, including Excel will have ways to limit the kinds of values you can enter into your dataset.

Some examples:

- Variables that should be numbers should only allow you to enter numbers.
- Categorical variables that are recorded with text should only allow you to enter predefined text for each category.
- Variables reflecting dates and times should only allow you to enter those using the exact same format. Again, some constraint on the range of acceptable dates or times can be helpful.

A Few Don’ts!

Wellcome - HRB
Clinical Research Facility
at St. James’s Hospital

- Avoid highlighting / colours etc. Keep it simple
- Avoid free text.
- Don’t include calculations in your spreadsheet NB!
- (DON’T USE EXCEL! Use REDCAP or similar instead)
  - excel lacks built-in security feature, can induce errors automatically, issues with version control, …..

Preparing for the establishment of health data access body services in Ireland under the European Health Data Space Regulation: Guidance for Data Holders

Status: Published on 16 Apr 2025

DOWNLOAD DOCUMENT

The European Health Data Space (EHDS) is the first common data space to emerge from the European data strategy. The EHDS Regulation entered into force in March 2025, with full implementation required between 2029 and 2031. The regulation aims to support the use of data for healthcare delivery (known as primary use of data) and to facilitate other uses of data, such as for research and policy-making (known as secondary use of data).

Under the regulation, data holders across Member States will be required to make certain categories of data available for secondary use, including electronic health record data such as medical history, diagnoses and treatments; administrative data such as hospital discharge records and research data such as clinical trials.

Further Resources re Datasets

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

- Data Organization in Spreadsheets (Karl W. Broman & Kara H. Woo)
- Sending me data in Excel (Kristian Brock)
- Tips for data entry in Excel (Crystal Lewis)
- Organize Excel Data before giving it to a Statistician (Piktochart)
- Long vs Wide formats (Stef van Buuren)

Management of Data
- Security, GDPR, backups

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

St. James's Hospital

- How do you manage your data at SJH?

- Training at SJH?

- Versioning of datasets?

- Backups?

Management of Data
- File Naming Convention

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

Do you ever create multiple datasets?

File naming at IDS-TILDA:

<parent file if applicable>_<description of file >_<creator initials>_<date of creation>_<version number>

For example: **wave 4 2021-03-31_data for JS_PP_16-02-2022_v0.1**

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The Role of Statistical Thinking and How to Get Help!

Data matters!
The Role of Statistics in Research

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

For high quality research, statistical thinking is required at all stages above - not just for a sample size calculation!

Sample Size Estimation

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

- We always want to conduct the smallest, most efficient but still effective trial studies (ethics, budget,…)
- Careful statistical thinking is required for sample size estimation
- ?What is a clinically meaningful effect of your treatment?
- ?How similar or different do you expect your trial participant outcomes to be? (across patient variability)
- ?How similar or different do you expect you participant measurements to be across time in each participant? (within patient variability)
- ?Power – how strong do you want your metal detector to be?
- ?Statistical significance?
- Do ask a statistician for help!

CAUTION!
Avoid Research Waste

- Most publications include results from statistical analyses
- High level of statistical error noted in several reviews of medical journal articles
- Gaps exist in researchers' statistical knowledge
- Researchers misunderstand / ignore statistics make misleading inferences / claims.

thebmj Research Education News & Views Campaigns Jobs

Feature » Essay

Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers

BMJ 2018 ; 363 doi: https://doi.org/10.1136/bmj.k4645 (Published 12 November 2018)
Cite this as: *BMJ* 2018;363:k4645

Linked opinion
Is 85% of health research really "wasted"?

Linked opinion
Funders and regulators are more important than journals in fixing the waste in research

Article Related content Metrics Responses

Paul Glasziou, director [1], *Iain Chalmers, coordinator* [2]

Author affiliations

Correspondence to: P Glasziou Paul_Glasziou@bond.edu.au

Progress has been made towards reducing the 85% of wasted effort in medical research—and the huge amounts of money misspent and harm caused to patients—but there's still a long way to go, say **Paul Glasziou** and **Iain Chalmers**

In their history of the evolution of guidelines for reporting medical research, Doug Altman and Iveta Simera showed that poor design, conduct, and reporting of medical research have been concerns for over a century: "The quality of published papers is a fair reflection of the deficiencies of what is still the common type of clinical evidence. A little thought suffices to show that the greater part cannot be taken as serious evidence at all."[1]

How to get Help?!?

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

DISCUSSION

- To produce quality research, you need statistical training and help with data and statistics!
- What available supports are there?
- How can I help? Monthly/ Quarterly CRF stats support clinic?
- Using ChatGPT / AI with due care!

HSE statisticians – grades approved in 2023

HSE our health service

Services | Health A to Z

Home > Jobs > **Eligibility Criteria**

- Medical Consultant Hub
- Recruitment Process
- Recruitment Licence
- **Eligibility Criteria**
- Validation of Qualifications

Statisticians

Statistician (Grade Code 0380)

Statistician, Senior (Grade Code 0381)

Statistician, Principal (Grade Code 0382)

Statistician, Chief (Grade Code 0383)

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

INTERACTIVE DISCUSSION

What to Collect / Not Collect (Meaningful Data)?

What to do with the data I collect?

What to Collect / Not Collect (Meaningful Data)?

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

DISCUSSION

- What are you collecting now and why?
- What is your research question? Objectives?
- Are you curious about some aspect of the data or the topic?
- How is the data that you are collecting going to answer your research questions / curiosities?

ADVICE

- Have clear objectives, discuss these / brainstorm with your colleagues!
- Make a plan in advance.
- If you had a magic wand that could generate this data now, what would you do with it? Would it be able to answer your research question?
- Is there any data already existing that can answer your research question?

What To Do With The Data You Collect?

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

DISCUSSION

- What can you do with the data?
- What software do you have access to?

ADVICE

- Become comfortable with the software that you have, strengths and limitations.
- Sketch your ideas of visualisations that would answer your question / tell an interesting story.
- Statistical analysis AFTER a lot of summaries and graphs.
- Can you find journal articles on similar topics that you're interested in? How do they present their data? Summary tables? Graphs? Can you apply these approaches to the dataset you are collecting?

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Experimental vs Observational Research

Why Should YOU Care About Research?

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

- Clinical research contributes to patient care and best practice
- Evidence based medicine
  - "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" – David Sackett
- Avoid the Evidence – practice gap –
  - The delay in integrating high quality research findings into clinical decision making

Experimental vs Observational Studies

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

Experimental studies evaluate an intervention / treatment with a comparison (control) group that are untreated or treated as usual.

Observational studies do not assign treatment. Instead they record patient data, treatments, outcomes etc. as they occur in routine care

- Examples - Registries, case control studies, cross-sectional study, retrospective and prospective cohort studies

Observational studies can be used to gather information and generate hypotheses

- (It's not always ethical to do an RCT e.g. Does smoking cause lung cancer.)

Why Should YOU Care About Clinical Trials?

Wellcome - HRB
Clinical Research Facility
at St. James's Hospital

- Two treatments seem effective in treating your patients. Which is actually better?
- RCTs provide GOLD standard evidence for
  - improved patient care
  - improved patient outcomes
  - Informing treatment guidelines

  Examples

  - testing a new / altered treatment, testing a vaccine, testing a scan or blood test, testing patient management strategies

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Thank You For Your Attention!

Dr Caitríona Ryan;

Associate Professor in Statistics and Trial Design, Trinity School of Medicine

ryanc86@tcd.ie;