IA4PH 25/10/2024

Synthetic data for microsimulation of cancer

Maikol Diasparra, PhD
Microsimulation Analyst
Health Analysis Division
Statistics Canada

Introduction to Microsimulation

Definition
Microsimulation involves modeling individual units (e.g., patients) to predict outcomes.

Use in Healthcare
It's applied to simulate disease progression, treatment outcomes, and health economics.

Why It Matters
Allows for detailed and personalized predictions to aid healthcare decision-making.

Why Use Synthetic Data in Microsimulation?

Definition: Synthetic data is artificially generated data that mimics real-world datasets without exposing real patient information.

Advantages:

- **Data Privacy**: Protects sensitive patient information.
- **Data Availability**: Fills gaps when real data is scarce or incomplete.
- **Scenario Testing**: Allows for the exploration of hypothetical or extreme cases without ethical concerns.

Generating Synthetic Data for Microsimulation

Process:

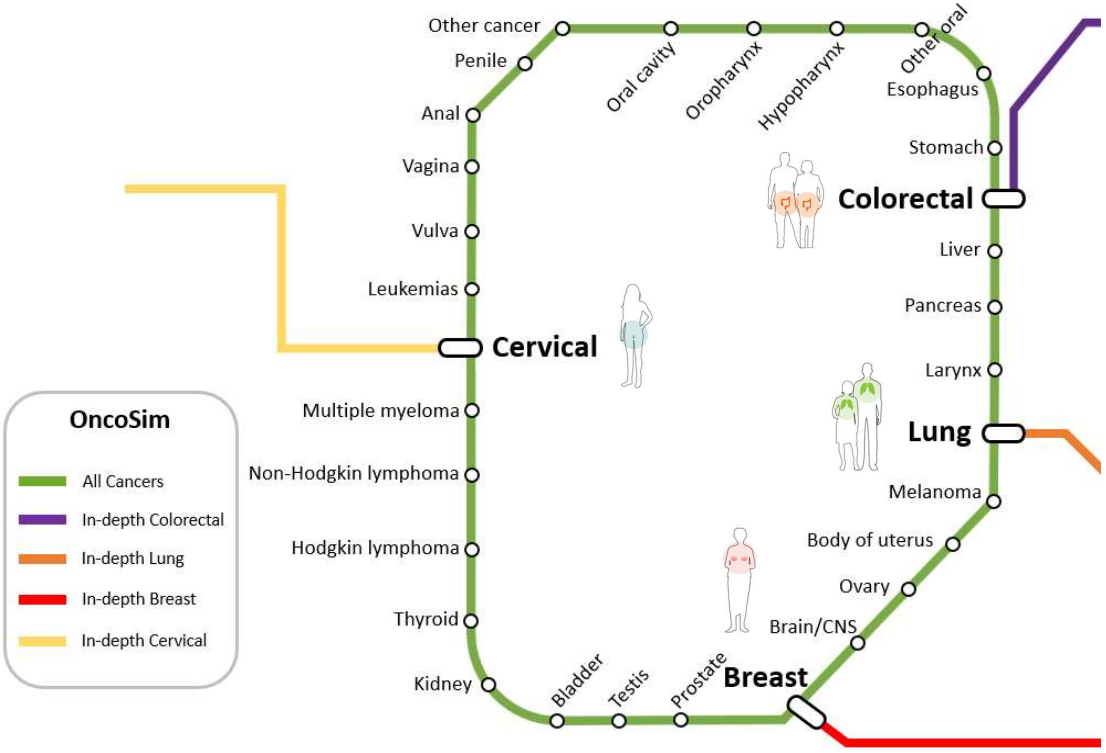
1. **Model Development**: Define rules and distributions that represent real-world cancer data.
2. **Data Simulation**: Use algorithms to generate synthetic datasets that reflect real-world characteristics.
3. **Validation**: Ensure that synthetic data mirrors real data in terms of key metrics like incidence and outcomes.

Challenges and Considerations:

- **Computational Resources, Bias and Uncertainty**

OncoSim

- Free, web-based cancer microsimulation tool
- Co-developed by CPAC and Statistics Canada
- Models breast, colorectal, lung and cervical cancers and related screening programs in detail
- Also provides high-level projections for 30+ cancer sites
- Projects health and economic outcomes and attributes them to 19 risk factors



How OncoSim works

Data

Inputs represent Canadian population and reflect disease progression, treatment pathways and costs in Canada.

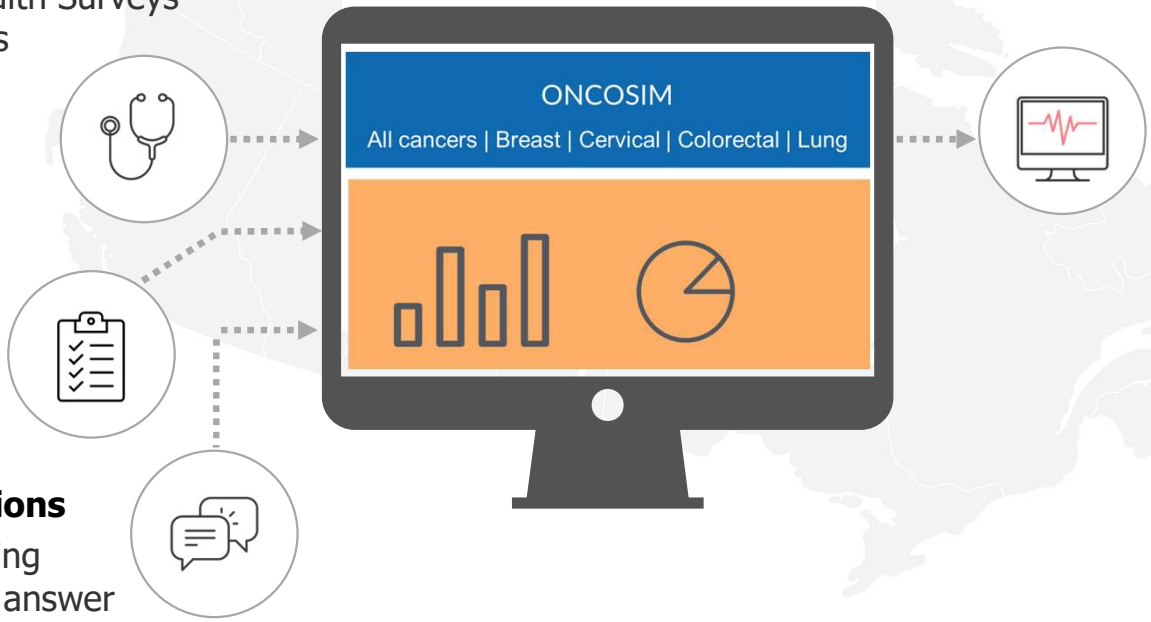
Sources:
-Canadian Cancer Registry,
-Canadian Community Health Surveys
-Official Demography Stats
-Healthcare admin data
-Clinical trial results
-Expert opinion

Verification process

-Calibration
-Internal and external validation
-Cross-model validation

Customizable assumptions

Users can modify the existing inputs and assumptions to answer questions.



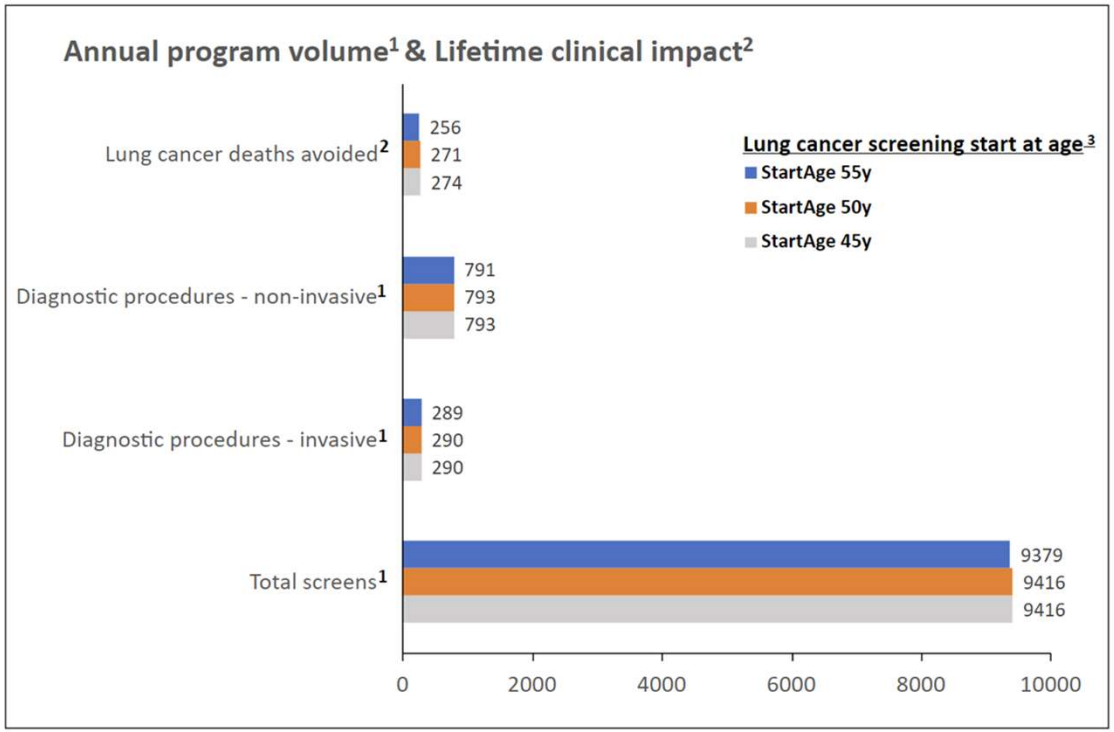
Insights

OncoSim projects provincial and national-level estimates:
E.g., # eligible for screening,
screening tests
colonoscopies
biopsies
cancer incidence, mortality, costs, etc.

Example 1: Lung cancer screening

Program Volume

- Starting lung cancer screening at 50 or younger requires few additional screening resources because few additional individuals meet the eligibility criteria.
- Starting earlier could save more lives by encouraging earlier smoking cessation, contingent on quit rates among those who are ineligible for screening but are still referred to smoking cessation intervention.



Note:

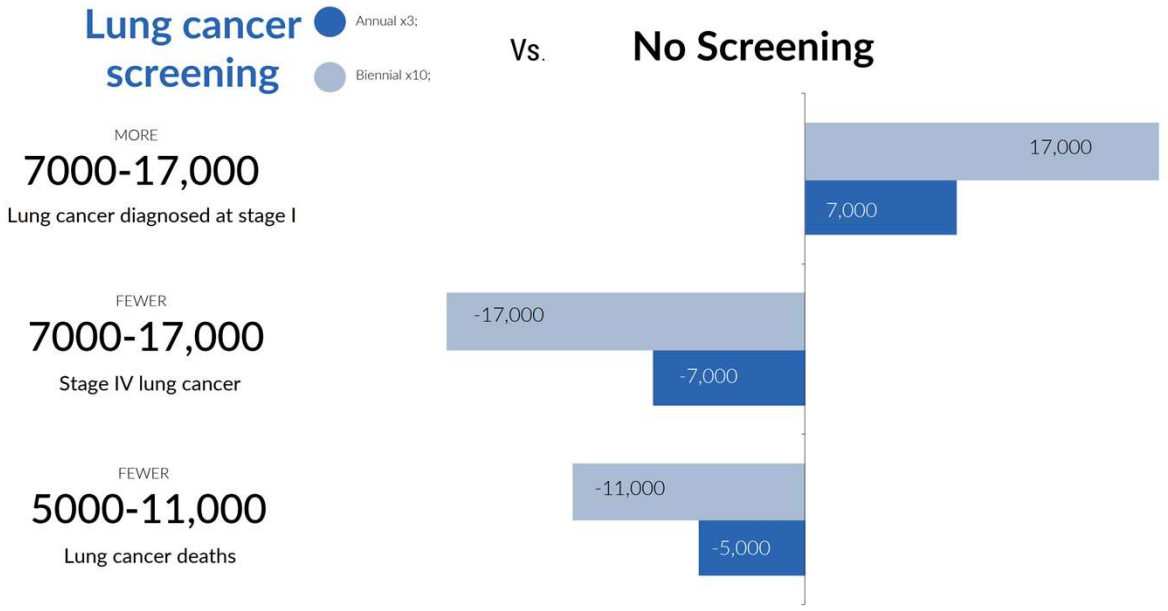
1. Annual average over the next 10 years, 2024-2033. Assumptions: 40% recruitment rate, 70% rescreen rate, and 5-year phase-in period.
2. Lifetime outcomes of a cohort born in 1950-1998, Screening vs. No Screening
3. Using 1.5% in 6 years risk threshold (PLCO-2L risk calculator)

Example 2:

Clinical impact of LDCT screening in Canada

Lung cancer screening could detect lung cancer earlier and prevent lung cancer deaths.

Compared to no screening, lung cancer screening in high-risk individuals cost <$25,000 per QALY gained and was cost-effective.*



*At the conventional cost-effectiveness threshold of $50,000 per QALY gained.

Conclusion

- Synthetic data plays a crucial role in enhancing the flexibility and privacy of microsimulation models.
- It enables researchers and policymakers to explore a wide range of healthcare scenarios without compromising real patient data.
- Future directions for synthetic data in cancer modelling include the use of advanced algorithms to generate more sophisticated synthetic datasets.

Getting involved?

If you think OncoSim has potential to support your work in cancer control, contact CPAC to get access

- oncosim@partnershipagainstcancer.ca
- https://www.partnershipagainstcancer.ca/tools/oncosim/

OncoSim is led and supported by the Canadian Partnership Against Cancer, with model development by Statistics Canada, and is made possible through funding by Health Canada.

CANADIAN PARTNERSHIP AGAINST CANCER | PARTENARIAT CANADIEN CONTRE LE CANCER

Statistics Canada | Statistique Canada

Thanks!

maikol.diasparra@statcan.gc.ca