LRES

LAND RESTITUTION EVALUATION STUDY

Impact Evaluation Report

March 31, 2023

Malcolm Keswell

Timothy Brophy

Patricia Chirwa

Kim Ingle

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

SALDRU
Southern Africa Labour and Development Research Unit

Land Restitution Evaluation Study
Impact Evaluation Report

Malcolm Keswell, Tim Brophy, Patricia Chirwa and Kim Ingle

School of Economics, University of Cape Town, Rondebosch, 7700, Postal Address: Private Bag Rondebosch 7700, Cape Town

31 March 2023

Implementing agency, funders and other government partners

Implementing agency

Funders

Government partners

Contributions

Malcolm Keswell was the lead-principal investigator on this project. He supervised all analyses, did the main impact analysis for the report, produced all of the tables and graphs and wrote the paper. Timothy Brophy oversaw all aspects of operations, data collection, data quality control, and data production. He also led the collaboration between the external fieldwork partners. Patricia Chirwa and Kim Ingle analysed data, prepared figures and provided other data support contributions to the study. Principal investigator Michael R. Carter was part of the team that worked on the initial design of the study and funding proposal. Mo Allouch provided technical assitance during the early stages of the study. Principal Investigator Mvuselelo Ngcoya provided inputs to the companion case study report.

Attribution

Please cite the work as follows:
Keswell, Malcolm, Timothy Brophy, Patricia Chirwa, Kim Ingle. 2023. *Land Restitution Evaluation Study: Impact Evaluation Report*. SALDRU, School of Economics, University of Cape Town.

Translations

If you create a translation of this work, please add the following disclaimer along with the attribution: *This translation was not created by SALDRU and should not be considered an official SALDRU translation. SALDRU shall not be liable for any content or error in this translation.*

Adaptations

If you create an adaptation of this work, please add the following disclaimer along with the attribution: *This is an adaptation of an original work by SALDRU. Views and opinions expressed in the adaptation are the sole responsibility of the author or authors of the adaptation and are not endorsed by SALDRU.*

Contents

List of Tables

List of Figures

Acknowledgements

From the implementing partner, the Commission on Restitution of Land Rights, we acknowledge the support provided by Chief Commissioner Nomfundo Ntloko-Gobodo, the Acting Deputy Land Claims Commissioner, Cindy Benyane, the Regional Land Claims Commissioner, Lebjane Maphutha, the Chief Director of the Western Cape Regional Land Claims Commission (RLCC), Wayne Alexander, as well as their Director for Quality Assurance, Rikus Janse Van Rensburg. We thank the Chief Director of the Eastern Cape RLCC, Zama Mamela both for his enthusiasm for the study and the support provided by his office. From the KwaZulu-Natal RLCC, we thank Chief Director Bheki Mbili, as well as Directors Walter Silaule, and Ndoda Mdluli. From the Gauteng RLCC, we thank Acting Chief Director, Mkhacani Makamu and Director for Quality Assurance Ilse Hayward. We are also incredibly grateful for the support provided by the Commission's Information Management Unit, especially Moraka Maredi and Maropene Mampjedi. Director of Finance and Supply Chain Management Francis McMenamin and Assistant Director of Supply Chain Management, Thabo Makhuto played pivotal contract management roles. We also thank Mangalane Du Toit and Sunjay Singh.

From the government-wide steering committee, we thank current and former chairs Godfrey Mashamba and Zakhele Mdlalose, both of the DPME in the Presidency. We also thank other DPME steering committee members, especially Thokozile Molaiwa, David Makhado, Jeanette Sprinkhuizen, Harsha Dayal, Phendukani Hlatshwayo, and Ahn-Lynn Poniapppen. From the National Treasury, we thank Bothwell Deka and Lebogang Madiba as well Molatela Masimine of DALRRD. Finally, from the University of Cape Town (UCT), we thank SALDRU's Director Murray Leibbrandt, who represented the University on the steering committee.

We also thank the international funder, 3ie, represented by their former Director Beryl Leach, both for her role as a member of the steering committee, as well as the formal role she later played as external reviewer of the final study reports.

A large team of UCT people were involved in the study, some not in any direct capacity. We thank Justine Burns, who was Director of the School of Economics at UCT for much of the time when the study was run. More recently, she provided detailed comments on earlier drafts of this report which were very helpful. Many other people were much more directly involved in the daily running of the study from fieldwork, to data production, to operations and we wish to recognise those contributions, especially the fieldwork coordinating team of Zuri Adonis, Alansa Klaasen, Lulama Mabaleka, Theophilus Mathoma and Nobubele Tyembile. Victoria Basapo, Michelle Bodzo and Morne Hoffman assisted with data quality control. Tania Hendricks and Amy Jephthah managed the study finances and HR.

Last, but not least, we wish to thank the Director of SALDRU, Murray Leibbrandt, who played a pivotal role in managing internal relationships between the study and other senior University stakeholders. The study has navigated a long road to completion and a hard won set of findings. Without the coordinated efforts of all stakeholders, this would not have been possible.

Abbreviations and Definitions

The Commission: "Commission on Restitution of Land Rights."

DALRRD: "Department of Agriculture, Land Reform and Rural Development."

3ie: "International Initiative for Impact Evaluation."

SALDRU: "Southern Africa Labour and Development Research Unit." It is a research division of the School of Economics at the University of Cape Town that carries out research in applied empirical microeconomics with an emphasis on labour issues, development, human capital, poverty, inequality and social policy. SALDRU was appointed by the International Initiative for Impact Evaluation (3ie) and the Department of Agriculture, Land Reform and Rural Development (DALRRD) to conduct this impact evaluation.

LRES: "Land Restitution Evaluation Study."

Restitution Act: "Restitution of Land Rights Act of 1994" (and its various amendments).

CAPI: "Computer Assisted Personal Interview." It is a generic term for the software programme used to administer face-to-face interviews.

CAPA "Computer Assisted Psychological Assessment." It is a generic term we use to refer to the software programme that we developed to administer the cognitive tasks to assess psychological well being and cognition.

Claim: the collection of individuals and/or their descendants that were dispossessed of property or tenure rights and who have submitted a claim for the restoration of those rights. There can be multiple claims originating from any given dispossessed property; for instance, a large community that was displaced as a result of apartheid era spatial planning.

CSM: "Core Sample Member." A CSM is a direct land restitution beneficiary that receives compensation. CSMs are tracked and interviewed if they are not found to be at their stated addresses.

NCSM: "Non-core Sample Member." Every household contains at least one CSM. A NCSM is a household member that is not a CSM.

Household: a collection of individuals that usually reside in the same dwelling unit, compound, homestead or plot of land at least 4 nights a week, including young children. In general there are multiple beneficiary households per claim.

ID Number: the number indicated in a CSM's South African identification document. This number is never visible to field teams but is made available by the Commission to the core team of authors of this report in order to ascertain the values of restitution awards.

Tracking: the process of locating CSMs with an incorrect address.

Tracing: the process of locating CSMs with no listed address through their ID numbers.

Claims Register: administrative data on the universe of claims at various stages of the restitution approvals process.

S42D: a legal agreement that is prepared in accordance with Section 42 D of the Restitution Act. Every claim requires a S42D to be prepared that details all aspects of the claim; for instance the location of dispossessed property, the values of restitution awards (cash to be paid or value of land/s to be restored) etc.

Settlement: the S42D agreement has been signed by the designated signatories of the Commission. A signed S42D indicates that the claim is considered approved ("Settled") in the legal sense.

Commitment Register: a ledger system that logs the commitment of funds for finalising a claim as well as the dates of settlement and finalisation.

Finalisation/finalised: a claim where the transfer of title deeds and/or cash disbursements has been completed.

Control group: sampled claims that are settled but not finalised.

Treatment group: sampled claims that are finalised.

Section 1

Introduction

1.1 Policy context

This study seeks to evaluate South Africa's Land Restitution programme. This programme has been implemented, at scale, since the democratic transition 25 years ago. The political, economic, historical and cultural backdrop of this programme is well known globally and is chronicled in a trove of scholarly and other works. There is a vast collection of ethnographic studies that have sought to document the brutality of land confiscations as well as many well-documented cases of the destruction of entire communities. In much of this ethnographic work, the devastation experienced by the victims often centres around the impact on family structures and the deep psychological scars left in the wake of the destruction. The basic facts are that between 1913 and 1994, millions of non-whites were displaced in South Africa. Through a process of evictions and forced relocations, much of the fertile and productive land came under the control of white South Africans, leaving the black population with marginal and relatively unproductive land. This was not restricted to rural agricultural regions; in urban areas evictions sought to homogenise neighbourhoods and often moved blacks to the periphery of the cities. It is estimated that at the height of the Apartheid years (1960-1983), over 3 million (out of an estimated 20-24 million) non-white South Africans were removed from their lands by the state. Furthermore, when including the removals due to other racially charged land consolidation policies, the number of dispossessed is estimated to be over 7 million people [1].

1.2 Evaluation questions

The land Restitution programme has been implemented, at scale, since the democratic transition 29 years ago. The Restitution of Land Rights Act of 1994 (and its various amendments) as well as Section 25, paragraph 7 of the Bill of Rights of the Constitution, are the two main pieces of legislation that mandates land restitution in South Africa. These laws seek to reverse the discriminatory policies of white land consolidation under the apartheid era such as the 1913 Natives Land Act, the Group Areas Act of 1950 and other related legislation of the Apartheid era such as the Black Authorities Act of 1951. The Restitution Act, provides for the restitution of land or other forms of compensation (generically defined in Section 25.7 of the constitution as "equitable redress"). In the main, equitable redress takes the form of financial compensation. Restitution, whether in the form of land or financial compensation, can be made to persons or entire communities, depending on the nature of the historical dispossession. The resulting transfers vary in size depending on the details of whether the claimants are family members or communities, and how the rights of succession have been applied to the surviving descendants of the originally dispossessed individual, but the transfers can be, and often are large. **The overarching evaluation question therefore, is whether such once-off transfers can impact overall individual and household welfare. Our focus in this report is on three core impact domains: economic well-being, psychological well-being and cognitive capacity.** In Section 1.4 of this report, we outline the rationale that motivates our focus on these impact domains.

1.3 Political economy context

Land reform is central in the public discourse over economic transformation and self-determination in South Africa. However progress towards achieving this objective has been slow and has become a flashpoint in the public discourse. In the context of the land restitution programme, the fact that many claimants "choose" not to be compensated with land is a further flashpoint, and stakeholders often attribute this tendency to something other than the role of choice. For example, many (perhaps most) claims involving land transfers entail far greater complexities and potentially longer waiting periods. Claims that involve very large communities often cannot be "settled" (both in the legal sense as well as the physical sense) where every beneficiary is relocated at the same time. One can think of many explanations about why so few beneficiaries select land (as opposed to financial compensation) that could be related to both administrative challenges but also to the "preferences" of the beneficiaries themselves. Preferences here could relate to the direct wishes of the beneficiaries themselves, but also to their indirect preferences in the sense of being patient enough to endure the much more complicated and time consuming processes of choosing to receive land.

The fact that land ownership is a flashpoint in the political discourse cannot be separated from the overarching problem of low social cohesion in South Africa. The process by which claims are settled requires active citizenry: the beneficiaries collectively must decide on what outcomes they prefer. While the intent here clearly is to engender agency and participation, it is at odds with the reality that this aspect of the process cannot work effectively without social cohesion as a pre-requisite. If the people that register a land claim have little to no connection with one another, this by itself could potentially foreclose on the sort of protracted engagements that are necessary to decide where to move to, if land were to be selected instead of cash. Social connectedness is therefore likely to be a very relevant institutional factor in the context of land restitution implementation.

The degree to which a group of people are socially connected with each other can reinforce beliefs that are commonly held or transmit novel information that could change the dynamics of the interaction and potentially alter outcomes. In the current context, the connections among beneficiaries in any given claim is also a product of history and possibly historical accident. While apartheid era spatial planning was exacting in how it separated communities by race, it was random in how it went about relocating those people that were forcefully removed, as long as the racial demarcation of the destination aligned with the racial classification of the victim. This process was probably haphazard and uncoordinated in some instances and much more tightly controlled in other instances. If true, this would mean that forced removals had the potential to disrupt not only community structure but also family structure. Thus, we should expect that where displacements operated inside as well as outside of households (as would be the case where previously co-residing siblings were relocated to different settlements), the resulting levels of connectedness between beneficiaries (who are mostly second or third generation descendants of the originally dispossessed individual (ODI)) is likely to be quite low, especially if they retained few of their social ties from the past.

These aspects of the restitution programme necessarily require a level of abstraction in this report. However, we take up these issues in our companion report, which analyses the local political economy context within restitution claims. The purpose of that report is to shed light on what happens when beneficiaries opt for land restoration in place of, or in addition to financial compensation.

1.4 Evaluation framework

The esteemed African American literary critic Ralph Ellison, described life in Harlem in the following way:[1]

> To live in Harlem is to dwell in the very bowels of the city; [...] Harlem is a ruin – many of its ordinary aspects (its crimes, its casual violence, its crumbling buildings with littered areaways, ill-smelling halls and vermin infested rooms) are indistinguishable from the distorted images that appear in dreams, and which, like muggers in a lonely hall, quiver in the waking mind with hidden and threatening significance.

But Ellison goes on to say,

> [I]f Harlem is the scene of the folk Negro's death agony, it is also the setting of his transcendence. Here it is possible for talented youths to leap through the development of decades in a brief twenty years, while beside them white-haired adults crawl in the feudal darkness of their childhood. Here a former cotton picker develops the sensitive hands of a surgeon, and men whose grandparents still believe in magic prepare optimistically to become atomic scientists. [...] It explains the nature of a world so fluid and shifting that often within the mind the real and the unreal merge, and the marvellous beckons from behind the same sordid reality that denies its existence.

These reflections about Harlem are evocative for several reasons. The first quote underscores the fact that distressed neighbourhoods are poverty traps: poverty, and every bad thing associated with it, reproduces itself, such that the material, social and psychological make up of the neighbourhood never changes, regardless of who moves in and out. Since Ellison's lucid essay, programmes to de-ghetto the major cities in the USA, from Chicago, to Boston and New York were implemented. These programmes were closely studied and there now exists compelling evidence that they have had startling impacts [3–9]. By carefully studying the positive life trajectories of these poor families that moved out of chronically distressed neighbourhoods, and comparing their trajectories with families that remained behind, the negative causal effects of distressed neighbourhoods has been scientifically established.[2]

Segregated spatial planning under apartheid produced places that resemble Ellison's dystopian version of Harlem. But unlike the efforts to desegregate America's major cities, South Africa's land restitution programme is far less targeted, and for that reason extremely complex to evaluate. To see why, we must return to Ellison's more optimistic second quote that describes the ghetto giving rise to success stories. What we learn from Ellison's essay is that poverty begets poverty, and escape is possible, but extremely rare.

[1] The quotes are taken from an essay penned by Ellison 1948 and first published in 1964 [2].

[2] The best evidence for the USA comes from the Moving to Opportunity (MTO) experiment conducted between 1994 and 1998 in five large U.S. cities. Randomly selected families living in high-poverty public housing projects were offered vouchers to subsidise their rental costs as an incentive to get the families to move into low poverty neighbourhoods. These families were then tracked and interviewed for over two decades. The findings show striking long run impacts for the families, and especially the younger children of the beneficiary families. See also **https://opportunityinsights.org/paper/newmto/**

These lessons offer us a lens through which to view the target population of restitution. The people in this evaluation are in most cases, the surviving descendants of families (many of them young children) that underwent a reverse type of mobility experiment, where they were moved, by fiat, into neighbourhoods that, virtually overnight, took on characteristics of the bad Harlem. However, as we have learned from the qualitative case studies, documented in our companion report, there can be quite a wide variety of experiences where some families avoided these outcomes, by drawing on kinship and social networks to move to alternative spaces. It is quite likely therefore, that we should expect the material, social and psychological circumstances of beneficiaries to be spatially heterogeneous and anchored, in some sense, to the site of the original dispossession. On one end of the spectrum, the more dramatic the departure, the less likely it would be that restitution could be expected to produce any lasting change for the beneficiary. On the other end of the spectrum, where the departure is less dramatic, the greater is the likelihood of change.

The Ellison quotes, and the evidence concerning poverty traps, foregrounds the opportunities and constraints entailed in assessing the impact of restitution on poverty. On the one hand, since restitution is based on a principle of equitable redress, the benefits *could* represent sizeable once-off cash or land transfers. For beneficiaries that receive substantial transfers, the impacts could be large and lasting. Poverty becomes a trap when outcomes enter a downward spiral. For instance, very impoverished households are not able to save, because they live from pay check to pay check. A sizeable transfer of income, or an income generating asset, could help kick start a plan to save, thus making it possible for the longer-term living standard of the household to rise. One way to view this kind of dynamic, is simply to ask, all else equal, whether consumption in the present is higher than consumption in the past, controlling for the size of the transfer and inflation. If current consumption is lower than past consumption, then a poverty trap is at work. To break the trap, the transfer needs to be large enough to allow new possibilities. Every model of a poverty trap embodies the notion of some type of minimum threshold that must be reached. We do not need to look at the data to accept the idea that two equally poor families that receive different levels of a transfer (one "small" and one "large") face very different long run prospects. If poverty traps are at work, small transfers can't be expected to shift the needle. The theory on this point is clear [10–13].

This framework undergirds our approach to evaluating the land restitution programme. Because of the nature of the programme, very large transfers as well as very small transfers are made to beneficiaries. We therefore will have naturally occurring variation in transfer sizes in our sample. If beneficiaries are trapped in poverty, we should not expect all transfer sizes to result in the same type or magnitude of impact. But we also need not have negative priors either. There is a vast literature on the effectiveness of cash transfer programmes across the developing world [14–17]. A central focus in this work is on the positive human capital development for beneficiares, particularly that of the children of recipient households [18–20]. The evidence for asset transfers is equally compelling [21–23].

Much of this recent evidence on cash and asset transfers is in principle relevant to the study of restitution. However, of greater interest are studies that have focused on large magnitude transfers and here the evidence is quite compelling that sizeable transfers can make a big difference [22–24]. Compelling new evidence also shows knock on effects for markets and as well as multiplier effects that shift the living standards of whole communities [25, 26].

1.5 Scope

1.5.1 Consumption

Consumption is the bottom line outcome measure we focus on in this report. household consumption will change if a big once-off transfer has lasting impact. In the main, beneficiaries receive substantial once-off cash or land transfers. Both types of benefits could see lasting effects for the beneficiaries, especially if poverty traps are at work. In §1.4, we outlined the theoretical framework undergirding this focus. The benefits of land restitution (whether in the form of a large cash payment or a right to land) have the potential to interrupt poverty traps. One method of detecting whether such a process might be underway would be to test whether there is a significant improvement in living standards for beneficiaries post receipt of transfer. In §2.1 we describe how consumption profiles were measured.

1.5.2 Psychological well-being and cognition

Improvements in psychological well-being (happiness, life satisfaction, stress and depression) has been a major focus area of impact evaluations of cash and asset transfer programmes [24, 27, 28]. The population in this study are descended from families that suffered severe trauma. Psychological distress early in life can have several, potentially interlinked effects on long-term well being. There is a vast ethnographic literature (as reflected in our companion qualitative case study report) that documents the psychological effects under various forced relocations during apartheid [29].

While scientifically identifying the causal effects of parental trauma on children, later in life, is complex, the potential theoretical pathways are not difficult to imagine. First, the financial disruption caused by a forced relocation (loss of property value, loss of assets) is well-documented [1]. This part of the cost experienced by a given family – the purely financial cost – can negatively affect cognitive capacity directly [30, 31]. The secondary shock of stress can also be a channel. Parental psychological stress has been found to negatively affect parenting quality to younger children [32].

The financial disruption can also take a toll in other ways. For example, a state of sudden poverty can lead to increased negative emotion and stress which in turn can cause greater risk aversion and time discounting (impatience). These effects on preferences could then lead to a negative feedback loop causing poverty to become entrenched. One intermediating channel involved in this feedback loop, implicated in high risk aversion and time discounting, are compromised attention spans and the adoption of habitual instead of goal directed behaviours [33].

The reverse causation is also possible. Executive functions – the set of cognitive processes that are central to cognition, psychological inhibition (i.e., blocking out intrusive thoughts and worries), memory, problem solving and planning have been extensively studied [34–37]. A key finding in this literature is that working memory (short-term recall) is negatively affected by depression [38]. The long term effects of psychological distress have also been shown to lead to lower levels of psychological inhibition [39].

In the parlance of restorative justice, psychological healing (or psychological closure) is probably an important impact domain for the above reasons. It would be a central outcome domain that is directly relevant for any surviving victim of a forced removal, but the effects could also be indirectly relevant, potentially effect modifying, either positively or negatively, for the surviving children who receive restitution transfers.

In §2.2 and §2.3, we describe in detail how we have measured both psychological well-being and cognitive capacities.

1.5.3 Social cohesion

Trust, cooperation, and what might be called community efficacy – the cohesive functioning of a community through better social relations, attitudes, and civic cultures – are important impact domains. One important dimension of social cohesion is trust: communities with high levels of social cohesion exhibit higher levels of trust and trustworthiness than less cohesive communities. Another important feature are the levels of cooperation and civic engagement within communities. Communities that have higher social cohesion will be more actively engaged in its development and generally display less cynicism about the prospect of improvements to be made.

Although it is possible, even likely, that communities that were destroyed by forced removals had high levels of social cohesion, to what extent do newer social ties, and connectedness, fostered by the restitution programme, promote social cohesion? Children that were forcibly removed could have had greater exposure to *negative social capital* (e.g., gang culture) and are thus likely to exhibit less trusting and trustworthy behaviour compared to their older counterparts who would have had a lower probability of exposure to negative social capital (because being older would create fewer opportunities for recruitment into a gang). Therefore, one measure of social cohesion would be to estimate trust and civic engagement levels of older versus younger beneficiaries.

1.5.4 Effect modifiers

Improvements in household consumption potentially reflects many other changes too, that may or may not play a role in escaping a poverty trap. An effect modifier is an intermediating outcome domain. Later, we explain that to test for an effect modifier, we need to test for whether the transfer causes a change in some other outcome that is not household consumption, and the impact on consumption itself varies in whether the intermediating outcome changed as a result of the transfer. Theoretically there could be impacts on many other domains beyond those that have been measured, and many outcomes for which impacts have not been estimated may not also be effect modifiers. In this report, we focus specifically on the two impact domains as potential effect modifiers - psychological well being and cognition – for which measurement was possible.

Section 2

Measures

2.1 Consumption

The respondent answering questions about householdspending is asked about total household expenditure in the last 30 days for food and non-food items (see pp.29-33 of Technical Document 1 for the full list of items). These are summed to provide total food expenditure and total non-food expenditure, respectively. These two components are added to total rental expenditure to give aggregated total household expenditure. Because data collection took place over multiple calendar years, these expenditure data are deflated. Details of the indices used are given in the data documentation and Stata do files.

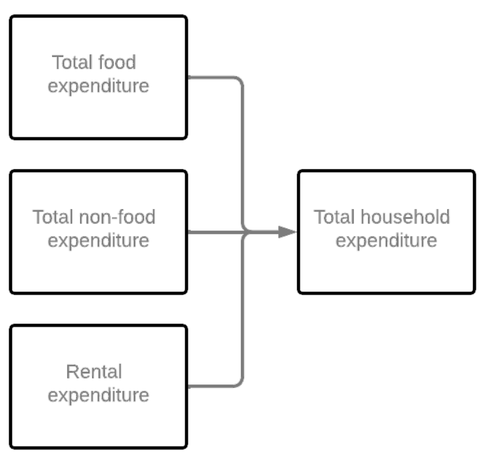


Figure 2.1: Construction of total household monthly consumption

Item non-response occurs when the respondent refuses to answer a question in the survey or states that they “Don’t Know” the answer. In these circumstances, imputation are performed on the individual variables affected. This is conducted only where a few qualifying conditions are satisfied. Single imputation regressions are run only when there are: (a) 100 or more “valid” responses for a variable; and (b) the percent of missing values do not exceed 40%. Pre-imputation, post-imputation and imputation flags are included in the individual derived and household derived data files for each variable that has been imputed.

Food expenditure is measured through a "one-shot" food expenditure question, rather than an aggregation of distinct food line items. We maintain the rule-of-thumb that imputation only takes place when there are at least 100 recorded observations and missing values do not exceed 40%.

If a respondent indicates that the household purchased one of the non-food items in the last 30 days, but cannot give an expenditure amount, this value is imputed using the same single regression imputation approach. Missing values for households that rent the dwelling household that they live in are imputed using a single imputation approach.

2.2 Psychological well-being

The population of beneficiaries under the land restitution programme have an important feature: the very act of historical land dispossession and the resulting poverty created by displacement is likely to have lasting effects on mental health. The key measure of psychological well being used in the study is constructed from the Centre for Epidemiological Studies 10 question depression inventory, or CES-D-10 (see, pp.193-196 of Technical Document 1) [40]. This version of the depression inventory has been validated and normed for South Africa [41].The CES-D-10 score indicates the extent of self-reported symptoms of depression experienced by the respondent in the past week. Each question is given a score from 0 to 3. If a respondent answers "Rarely or none of the time (less than 1 day)", their score for that question is 0, "Some or little of the time (1-2 days)" scores 1, "Occasionally or a moderate amount of time (3-4 days)" scores 2 and "All of the time (5-7 days)" scores 3. This rule applies to all questions other than 5 and 8, the two "positive" questions. For these questions, the score is reversed so that "All of the time (5-7 days)" scores 0, "Occasionally or a moderate amount of time (3-4 days)" scores 1, "Some or little of the time (1-2 days)" scores 2 and "Rarely or none of the time (less than 1 day)" scores 3. The total score is simply the sum of the scores for each of the 10 questions and will therefore be a score out of 30. If a respondent did not answer all 10 questions (i.e. they either refused or the answer is missing for one or more of the questions), their score is set to a missing.

2.3 Executive functioning

Executive functioning refers to the mental processes that dictate the way we absorb and interpret information such as working memory, inhibitory control, and cognitive flexibility. These capabilities are measured through four tasks that each interviewed household member completed.

2.3.1 Working memory

The Corsi- Block test measures visio-spatial working memory. The test is visual and is independent of background knowledge or ability. It has been shown to be the most appropriate measure and predictor of visual working memory capacity. Working memory is a key component of cognitive activity and is closely linked with the ability of an individual to control their attentional resources and perform on higher order cognitive tasks. The task is self-administered whereby a set of blocks flashes a different colour on a screen, one block at a time, in quick succession. The participant is then asked to recall the sequence, by tapping each block in the order that they changed colour. Several trials are run, where each successive trial increases in level of difficulty as the number of blocks that flash. Performance on the task is measured by the number of flashing blocks the participant is able to correctly identify. For instance, if in particular trial four blocks flash one at a time and the participant is able to correctly recall the order in which they flashed, then the score will be four. If on the next set of trials, the number of flashing blocks increased to five blocks and the participant is consistently unable to recall the correct sequence, then the subject's score for the task as a whole is four. This number is called the "span" and is the key measure of working memory discussed in this report.

However, working memory likely reflects more than just the capacity to recall information. It has long been recognised that attentional resources (or bandwidth) is involved in how well something is remembered. The ability to hold items in memory is a result of greater ability to control attention, not necessarily a larger memory store [42]. In order to capture this role of attention, we employ a further task to measure cognitive inhibitory control.

2.3.2 Inhibitory control

The Stroop task measures inhibitory control. A typical implementation is when participants are shown words of colour names such as "Red" and are asked to name the colour in which each word is written. The word and the colour of the word can be congruent or incongruent; e.g., "Red" can be written in the colour blue (incongruent) or red (congruent). An incorrect answer to an incongruent condition is aimed to measure cognitive control (attention). The task usually consists of many trials (30 in our case). Response speeds and error rates are recorded.

However, this version of the Stroop task is not suitable when literacy or language is an issue. We therefore implemented a version of the Stroop task based purely on images. Figure 2.2 shows an example of the task.

Participants are shown a picture of either a heart or a flower on a computer screen in rapid succession. The participant is asked to press on the same side of the screen on which a picture of heart flashes but to press on the opposite side of the screen on which a picture of a flower flashes. The participant has a limited amount of time to answer. The flashing flower is a form of proactive interference that is meant to trigger the need to resist the impulse to tap on the same side of the screen, when a flower appears after a sequence of hearts has appeared. This tests a person's inhibition, a component of executive function, used to suppress irrelevant stimuli. In the version of the task implemented in the study, the participant performs the task with 30 iterations each taking 3 seconds.

2.3.3 Fluid intelligence

The Ravens Progressive Matrix is a prominent and widely accepted pattern recognition task aimed at measuring logical thinking and novel problem solving [43] . It is non-verbal and designed to be independent of background knowledge or education. Fluid intelligence is important as it forms part of a person's ability to reason, solve problems and think logically. These capacities are known to be linked to working memory.

It is implemented by presenting the participant with the "puzzle" such as the one shown in Figure 2.3. The participant must then choose among the 8 choices the best fit into the shaded box in the lower right hand corner. In the below example, the answer is 6. The respondent is given 10 minutes to complete 10 such "puzzles". Each puzzle/trial gets progressively more difficult, such as the one shown in Figure 2.4. Performance on the task is measured by the percentage of correct matches.

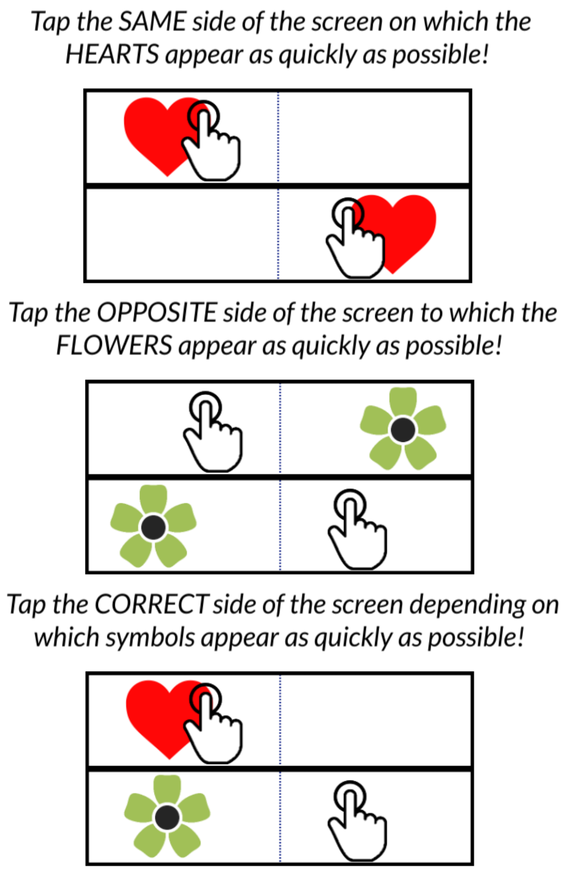


Figure 2.2: Stroop task example

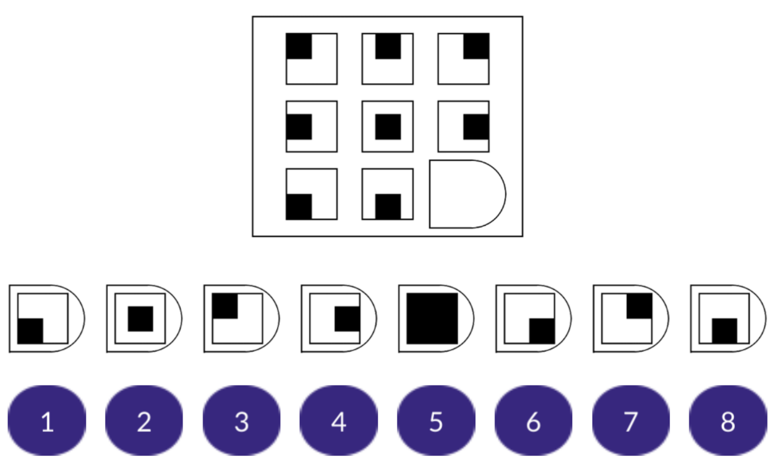


Figure 2.3: Raven's progressive matrix task: easy example

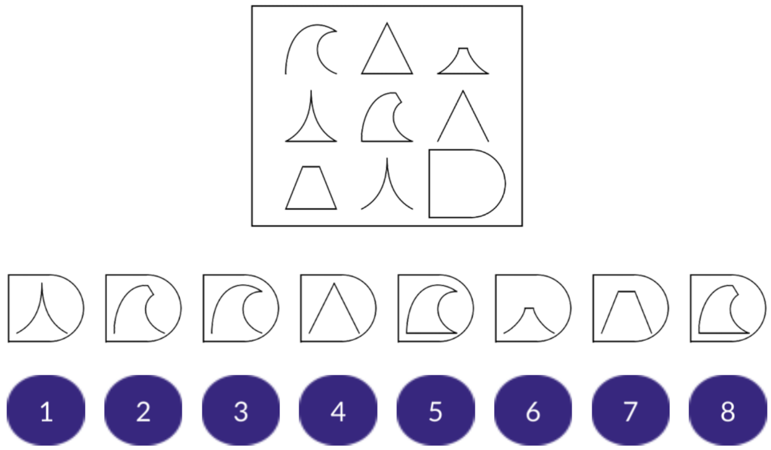


Figure 2.4: Raven's progressive matrix task: difficult example

2.3.4 Planning

The Tower of London task is an extensively used measure of planning ability. The task is administered by asking participants to mentally pre-plan their moves that would match a set of start discs to a goal. They then execute these moves one by one, with the objective of doing it as quickly and in as few moves as possible. Performance in this task is indicative of planning ability, a key component of cognitive control. In the example shown in Figure 2.5, the task can be completed in 4 steps: green to right peg, blue to middle peg, red to middle peg, green to left peg. The purpose is to measure whether the participant is able to resist the temptation to complete the task without first planning it out mentally. Generally, participants that complete the task (even if quickly) in more than the minimum number of moves will exhibit lower levels of executive function [44, 45].

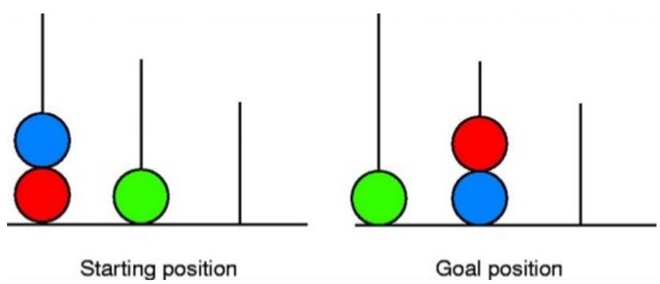


Figure 2.5: Tower of London example

Section 3

Methods

The complex set of processes that play out during the course of a claims approval poses a challenge for sampling and design. Since a claim can be dismissed during the verification stage, or can become tied up in the courts during negotiations with beneficiaries and landowners, the eligibility of a claim for selection into the study depends on whether it has successfully navigated all of these risks to approval. Section 3.1 deals with how we approached this important design choice.

Section 3.2 then outlines the more detailed steps we took to further screen claims and beneficiaries in order to build the sampling frame. In essence, we narrow down the set of eligible claims to only claims that have reached the settlement and/or finalisation stage. At this stage, the claim has been approved or settled in the legal sense. It then enters the finalisation stage where the award (either monetary compensation or restoration of land rights) is earmarked for transfer. The date of the final transfer thens depends on budgetary commitments. Generally a claim that has been settled/approved can be expected to be finalised within one or two fiscal years. For the control group, we included all claims that were settled but not finalised. The treatment group sampling frame is determined in a more straightforward manner; it is the universe of all finalised claims for the last ten years (2013-2022). A random sample (without untraceable replacements) is then drawn.

In Section 3.3 we then take up the second major design challenge: the sizes of the sample to draw from the sampling frame in order to estimate treatment effects sufficiently precisely. We present minimum detectable effect sizes for the planned sample, as well as the realised sample. The third and final issue discussed in this part of the report is the choice of statistical methodology. This is discussed in Section 3.4.

3.1 Determining claim eligibility for study sampling

The restitution programme entails an extremely complex process of legal verification as well as negotiations between the primary legislative body, the Commission for the Restitution of Land Rights, hereafter “the Commission,” as well as landowners, and intended beneficiaries. Figure C.1 of Appendix C shows shows the main stages of the vetting and validation of a land restitution claim.

Figure 3.1 narrows down the approval pipeline shown in Figure C.1 to the part that is relevant to the study.

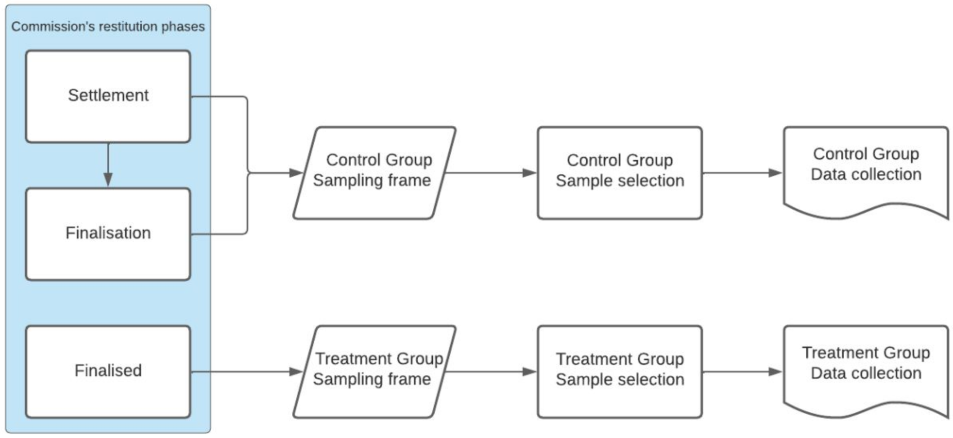


Figure 3.1: First step: narrowing the pipeline

3.2 Sampling

Restitution claims are generally associated with multiple beneficiary households because in the vast majority of cases a claimant will have submitted a claim on behalf of other individuals also affected by the original dispossession (for example, other adult children of the originally dispossessed individual). These features of the programme therefore necessitates a clustered design: claims are first randomly drawn, and then beneficiaries within a claim are randomly drawn. However, the clustered design does not require balance at the level of the clusters as none of the impacts we estimate are for whole claims.[1]

3.2.1 Constructing the sampling frame

The first step in the construction of the sampling frame is the collection and capturing of beneficiary data, verification of the quality of that data and finally the tracing of the beneficiaries themselves. We refer to this part of the process as Beneficiary data capture, Verification and Tracing (BVT). The culmination of the BVT process identifies the final sample frame, with untraceable beneficiaries being removed from the universe of captured claims and replaced by another randomly drawn claim. As already indicated in section 1, the design of the study is quasi-experimental (c.f. Figure 3.1). Selection bias therefore has to be dealt with statistically. This includes forms of selection bias that stem from untraceable beneficiaries, As we discuss in section 3, the methods we use indirectly also control for these type of sampling effects.

Figure 3.2 below shows the complete process flow. Below we elaborate on pertinent details involved in the BVT process (extent, timing, stage verification etc.).

[1] We do not look at claim level outcomes because claims don't necessarily relate to clusters in the standard sense of a community/village, or a unit of analysis that can only be defined over the cluster such as clinics, schools, etc., Some claims can represent "communities" but not necessarily spatially. Our companion case study report delves into these issues in greater detail by looking at patterns within claims that resemble communities in the sense of social networks.

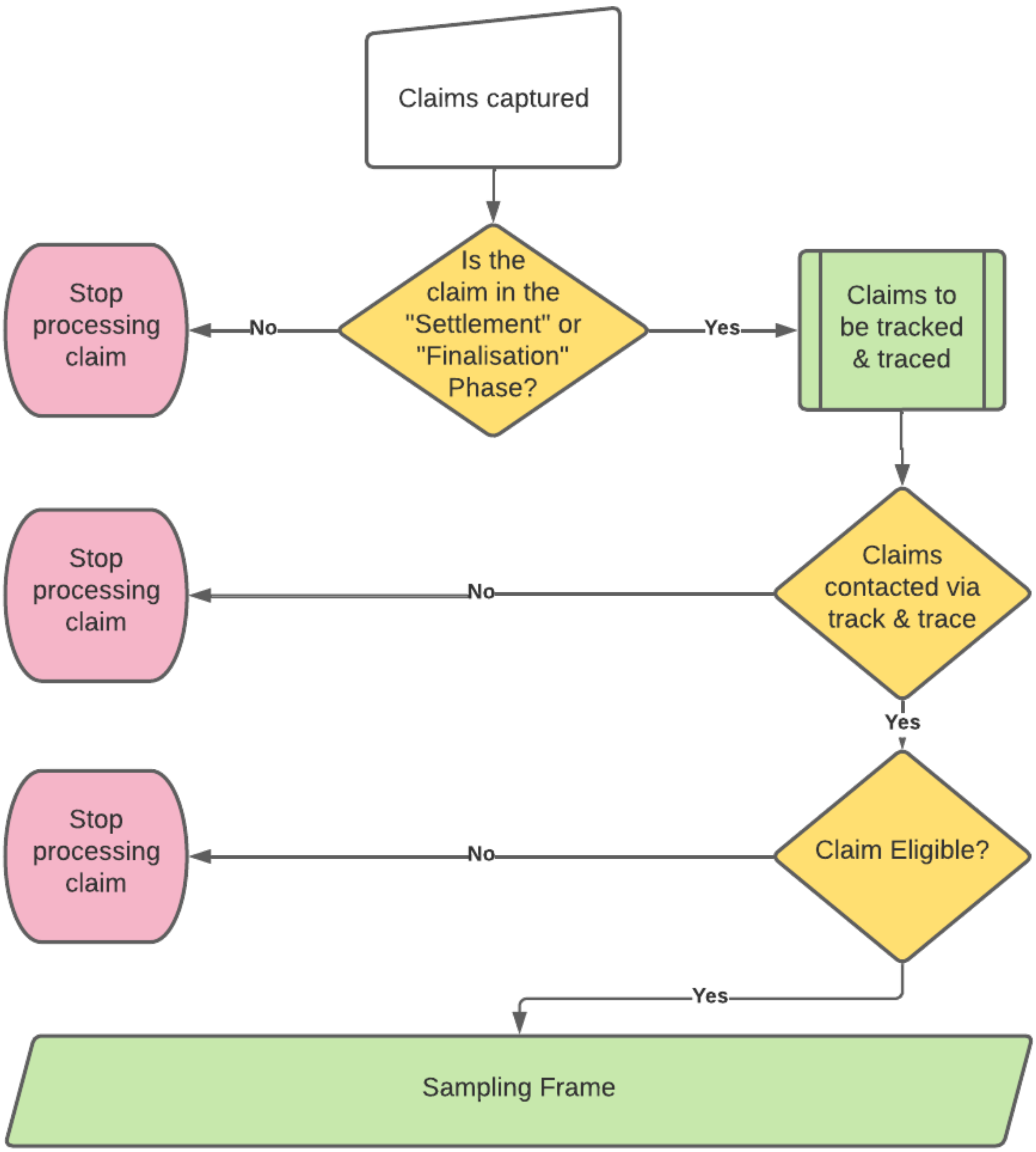


Figure 3.2: Constructing the sampling frame

3.2.1.1 Capturing of claim and beneficiary data

Capturing of the claim files started in 2017 and continued through 2019 starting with the development of software to capture the claims and then the capturing of the claims data from the physical claim files. In this first phase of fieldwork, we travelled to regional land claim offices in six provinces – Western Cape, Eastern Cape, KwaZulu Natal, Gauteng, Mpumalanga and Limpopo – to retrieve and capture claim documents.

3.2.1.2 Selection eligibility rule 1: determining the phase of a claim

Once the claims are captured, the eligibility for study selection is then determined. This step makes use of two administrative data sources – the Claims Register and the Commitment Register.[2,3] This is a particularly tricky part of the process as the Claims Register is an aggregation of data supplied by the regional Commission offices, whereas the Commitment Register is held and managed centrally. Since the regional and national offices often do not share a common claim file identifier, machine learning tools have to be deployed to link claims found in both registers. We used statistical string matching algorithms based on the claim name, extent of the land, lodgement date and property descriptions to merge the Claims Register data with the Commitment Register data in order to determine whether a claim had been approved (i.e., "Settled").

3.2.1.3 Track and trace the claim

Once a claim is determined to be "Settled" or in the "Finalisation" phase, the information that has been captured has to be verified as up to date, and the beneficiaries located. This Verification (V) and Tracing (T) part of the BVT process involves the following key steps:

- Sourcing of recent contact information for beneficiaries listed on the claim.
- Contacting the beneficiaries or their representatives.
- Confirmation that the beneficiary and their descendants are still alive.
- Confirmation that the claim has not been finalised yet.

When there is only one contact number for a whole group of beneficiaries (or the claim is being run as a community claim), permission is obtained from community leaders before proceeding with fieldwork.

[2] The Claims Register is a register of all claims in the pipeline. Claims in this list are at various stages of the restitution process so selecting claims from this list will have a low probability of being finalised within 12 months. Some claims will also eventually not be approved due to non-compliance with the legislation. Therefore a selection rule is applied to this register in order to ensure that only claims with a high probability of finalisation over the next 12 months are selected. The selection rule we applied is to check that the last completed milestone of a claim was approval of the S42D Agreement. This is an indicator that a claim has been "Settled."

[3] The Commitment Register is the key monitoring tool used by the Commission to cross check payments and finalisation status of projects. Once a project/claim has been settled, the amount to be paid out (to acquire and transfer land or to pay out cash compensations to beneficiaries) is recorded as a commitment in this spreadsheet. Therefore if a claim appears in this register, there is virtually zero probability that a claim will not be finalised within the fiscal year. There is also little chance it will be found to be non-compliant. If a claim that has been selected from the Claims Register has been finalised, it will appear on the Commitment Register. The Commitment Register also includes data on the beneficiary count. We use the Commitment Register to eliminate claims from the sample that appear as "Settled" in the Claims Register but have in fact been finalised (which means treatment has been effected)

3.2.1.4 Selection eligibility rule 2: determining if there is at least one eligible beneficiary

Once a claim has been contacted, we then determine if it is eligible to form part of the sampling frame. Eligibility in this sense is when the beneficiaries are not all deceased and we have been able to make contact with the claim or have made contact with individuals that could help in tracking the remaining beneficiaries.

The remaining claims that have not been eliminated by these eligibility criteria, defines the sampling frame used, both in Phase 1 (2018-2019) and Phase 2 (2021-2022).

3.2.2 Drawing the samples

There are three distinct (and necessary) stages to drawing a sample from the sampling frame. Figure 3.3 shows the decision tree involved in drawing the final sample.

3.2.2.1 First stage: selection of claims/clusters

The sampling frame is at the level of the claimant/land parcel. A settled claim equates to a land parcel to which many individuals might have restored rights. In the event that the group decides to opt for cash, then the cash value of the land will be divided equally among members of the group. Groups that opt to receive land will all have an equal stake in the returns generated by the land. Thus the land parcel/claim represents the primary sampling unit (PSU).

3.2.2.2 Second stage: setting the cluster size

The second stage sets the ideal sample size per claim (or cluster size). For family cash claims (i.e., claims comprised only of related individuals electing to receive cash), all beneficiaries in the claim are selected for interview. This is possible as these types of claims are smaller on average (7.1 beneficiaries per claim). However when dealing with land restoration claims (again only involving related individuals), the average number of beneficiaries per claim is much larger (30.3 beneficiaries per claim). Because the intra-cluster correlation (ICC) of outcomes is generally higher for smaller family claims, we set the number of target beneficiaries per claim to be on the higher side, randomly selecting a maximum of 40 distinct branches of the family tree for any given family claim. This rule in theory could mean 40 people from the same household are selected, but this is exceptionally rare.[4]

3.2.2.3 Third stage: selection of individuals within a unit

A household must contain at least one Core Sample Member (CSM). CSMs are individuals who are direct land restitution beneficiaries that will receive compensation. In Phase 2, only CSMs were selected.[5] However,

[4] The aim of this sampling algorithm is to maximise the likelihood of distinct units, even though all share a common ancestor. For instance, a family of 5 siblings with a deceased parent who was dispossessed will all have an equal right to restitution. If one of these siblings is deceased, then their rights pass to their children, and so on. With the average age of the originally dispossessed person being above 80, the age patterns of restitution dynasties tend to be generally on the older side, which means that it is likely that first and second generation descendants also have a high probability of being deceased. Therefore, even if one of the 5 siblings, in this example, is deceased, that person's rights could potentially transfer to tens or hundreds of other people depending on how large the extended family is across the many generations that define the dynasty. Setting the limit high enough then gives more weight to sampling such large dynasties that split into many distinct households as we move down the family tree to younger non-coresident beneficiaries who could be grandchildren or great-grandchildren.

[5] Many aspects of the survey instruments and systems changed following the COVID-19 pandemic. Questionnaires were shortened and moved to a Computer Assisted Telephonic Interview (CATI) platform. One of the necessary changes was to narrow the focus to only focus on direct beneficiaries in order to shorten the length of the interview.

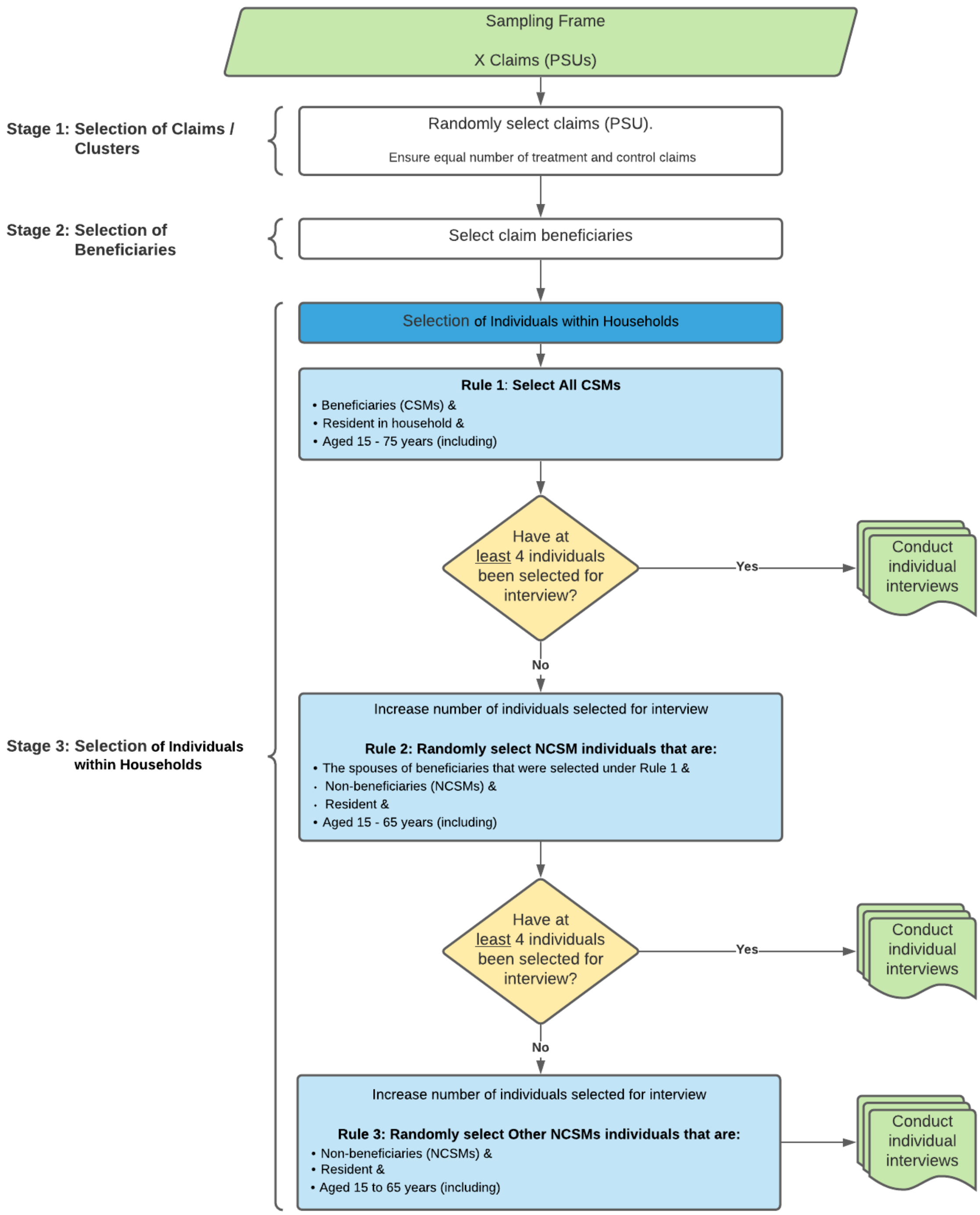


Figure 3.3: Sampling algorithm: phase 1

in Phase 1, if a household has fewer than 4 CSMs, Non-Core Sample Members (NCSMs) are also selected. The algorithm used to make this determination has three rules.

Rule 1 selects individual CMSs aged 15 to 75 years of age. This rule accounts for two features of the population of restitution beneficiaries. First, the fact that surviving originally dispossessed individual (ODI) will generally be older than 70. Second, the fact that many of the outcome domains under study covary with age (for instance, psychological well-being and cognitive functioning both decline naturally in older populations). Placing an age cap at age 75 mitigates the likelihood of this form of selection bias.

Rule 2, selects NCSM spouses of CMSs selected under rule 1. If the beneficiary does not have a spouse, or the spouse is deceased, Rule 3 then selects any other NCSM fulfilling Rule 1.

3.3 Sample sizes and impact detection

In this final section, we investigate minimum detectable effects (MDEs) for different sample size scenarios. MDEs are thresholds than can be calculated for outcome variables for which data exists. These thresholds are informative about the size and scope of a study as it gives us important information for how to benchmark expectations of what magnitude of impacts the study is capable of finding, should they exist.

This type of exercise was part of the design phase of the study in 2016 to demonstrate feasibility to the funders. However, due to the effect of the pandemic, major aspects of the study design and data collection had to be reassessed. This section of the report presents the new sample targets that we worked with during Phase 2 (post-pandemic). The resulting estimates differ from the prior ones, in two ways. First we updated the formulas to use the descriptive statistics from the Phase 1 sample.[6] Second, we updated the estimates with information for the whole sample.

3.3.1 Consumption

We begin by calculating MDEs for per-capita consumption, which is the main outcome variable for which we will estimate impacts. Table 3.1 shows a summary of the results. The MDE is an estimate of the smallest level of impact that the applicable design is capable of detecting. For instance, the value of 17.48 reported in the last column of row 5 of the table indicates that an impact of 17.48%, and no smaller, can be detected with a sample of $n = 7 \times 500 = 3500$ households, where the ICC is set to 0.20.

Table 3.1 also shows the effect of sampling one additional household per claim (from an expected realisation of 6 households per claim). The results indicate that adding another 348 or 500 households can be expected to reduce the MDE by a further $1.5 - 1\%$. There is therefore a negative relationship between MDES and increases in sample. Bigger samples make it possible to detect ever smaller impacts. This is what we would expect and hope for: if the impacts are small, we would want a large enough sample to detect it in the data.

However, there is a point at which trying to estimate an impact that is so small that it does not justify the cost it would incur to prove such a hypothesis. To illuminate the full extent of our benchmarking of desirable impacts and financially justifiable data collection, Figure 3.4 shows the relationship between targeted impact detection and samples per claim varying between 1 household per claim and 10 households per claim. It is clear from the graph that sampling additional claims (250 per treatment arm for a total of 500 claims) does

[6] In 2016, the necessary descriptive statistics (mean, standard deviation, intra-cluster correlation) had to be inferred from a prior study [22]. At the inception of this study no prior representative survey had been conducted on restitution beneficiaries so there was other alternative but to base our assumptions on related sample

Table 3.1: Target minimum detectable effects: per capita consumption (% change)

Intra-cluster correlation	358 claims		500 claims	
	n=6	n=7	n=6	n=7
0	19.96	18.56	16.97	15.79
0.05	20.43	19.08	17.37	16.23
0.1	20.88	19.59	17.76	16.66
0.15	21.33	20.08	18.14	17.08
0.2	21.77	20.56	18.52	17.48

n is the number of CSM households.Test size and power: we assume standard values of $\alpha = 0.05$ and $\beta = 0.8$. Mean per capita consumption for control group households estimated from Phase 1 of data collection is R3491.037, with a standard deviation of R2742.467. Two sample ranges are condsidered: a smaller sample of 6 or 7 CSM households each for 356 claims ($n = 2136 - 2492$); and a larger sample of 6 or 7 CSM households each for 500 claims ($n = 3000 - 3500$). ρ denotes the intra-cluster correlation (ICC). The ICC is a measure of the extent to which households within any given claim can be expected to have outcomes that are similar. For instance $\rho \approx 1$ means that every household in a claim has virtually the same level of consumption. By contrast $\rho \approx 0$ means that there is virtually no correlation between the consumption levels of households in a claim. We have allowed for a wide range of ICC values, with an upper limit of 0.20

indeed make a difference to the estimates. The reference lines in both panels of the graph give the plotted values for the 7- household designs that are reported in columns 3 and 5 of Table 3.1. The larger sample shown in the panel to the right can expect to deliver MDEs that are more than 4% smaller for an ICC of 0.20, compared to the smaller comparable sample for the same level of ICC.

As both the table and figure reveal, both smaller ICCs (the levels of the curves) and larger samples (the x-axis), deliver smaller MDEs. However, Figure 3.4 reveals something that is not apparent in Table 3.1: that the sharpest drops occur for relatively small numbers of households per claim. This is reflected in the slopes of the curves. Both panels of Figure 3.4 reveal that the curves are relatively flat after about 7 households per claim. It therefore would not be optimal to sample more than that many households per claim, as the gains, in terms of reductions in MDEs are not substantial relative to the escalation in cost.

Our preferred MDEs that we based the pre-pandemic design on was $13.07 - 18.68$ percent ($ICC = 0.2$). That estimate represents about half the level of impact reported in a related study [22], so the pre-pandemic design was extremely well powered.

The changes to the design that were necessitated by the pandemic changed this picture somewhat. Table 3.2 updates the ex-ante sampling target of 500 claims (for cluster sizes of 6 and 7) with the sample means, standard deviations and ICC of the final realised sample. The value of 24.24 reported in the last column of row 5 of the table indicates, as before, the smallest percentage impact that can be detected with a sample of $n = 6 \times 510 = 3060$ households, where the ICC is set to 0.15. The value corresponding to 510 claims and 5 households per claim is 24.75. This is the closest calibration we can make compared to the sample that was actually realised; 511 claims with a mean cluster size of 5.23 for a total of 2670 households with an ICC of 0.13. This figure is about 6% higher than both the upper bound of 18.68% estimated before the study began or the one estimated with the benefit of the data we collected before the pandemic (18.14% shown in Table 3.1). Given that the realised ICC of 0.13 for the full sample (Phase 1 and Phase 2 combined) turned out to be within the range of those used to recalibrate the sample sizes for Phase 2 (0-0.20), and that the target of 500 claims was exceeded (albeit with about 1 household less per claim), the main reason for

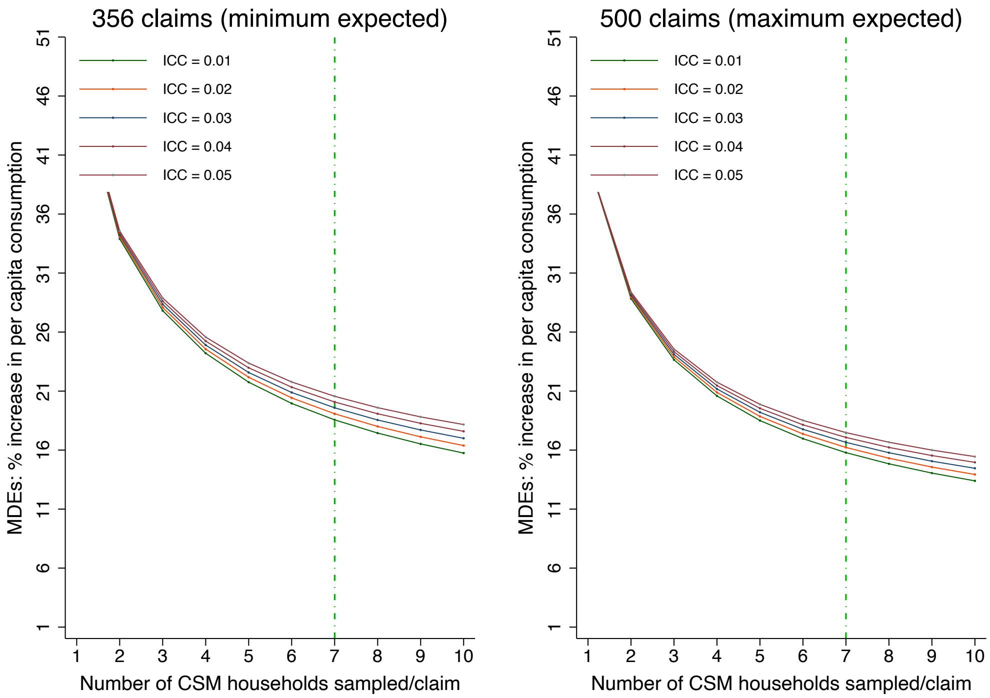


Figure 3.4: Minimum detectable effects for per capita monthly consumption

Test size and power: we assume standard values of $\alpha = 0.05$ and $\beta = 0.8$. Mean per capita consumption for control group households estimated from Phase 1 of data collection is R3491.037, with a standard deviation of R2742.467. Two sample ranges are considered: a smaller sample of $1 - 10$ CSM households each for 356 claims ($n = 356 - 3560$); and a larger sample of $1 - 10$ CSM households each for 500 claims ($n = 500 - 5000$). ICC denotes the intra-cluster correlation. Estimates represents minimum detectable effects (MDEs).

the somewhat higher MDE can only be attributed to differences in the distribution of the outcome variable for the full sample, versus the Phase 1 sample. For the full sample mean per capita consumption for the control group is slightly lower than the phase 1 control group of (R3229.48 versus R3491.037). However, the standard deviations are vastly different with the full sample being more than twice as variable than the Phase 1 sample (standard deviation of R6461.36 compared to R2742.47). This is the main reason for differences between the realised MDEs and the target MDEs.

This is not surprising in many ways. More than half the data collection was completed in 2022 some six months after pandemic restrictions were lifted. This fact alone might account for bigger spreads in the data than we would ordinarily see. It is therefore important to control for this structural break in the data which we do in all the estimated impacts presented in this report.

A final comment on the somewhat higher than expected MDEs: it turns out that an MDE of 24.75% was low enough after all. A simple, naive estimate of the treatment effect on real per capita consumption (log transformed) is 24.82% and is statistically significant at the 1% level. While we do not dwell on this naive estimate when discussing the results in the next section, it is reassuring that the size of the sample will not be an impediment to detecting average treatment effects on the treated.

Table 3.2: Realised minimum detectable effects: per capita consumption (% change)

Intra-cluster correlation	500 claims		510 claims	
	n=6	n=7	n=5	n=6
0	18.51	17.13	20.07	18.32
0.05	20.69	19.54	21.99	20.49
0.1	22.67	21.67	23.75	22.44
0.13	23.77	22.86	24.75	23.54
0.15	24.48	23.62	25.39	24.24

n is the number of CSM households. Test size and power: we assume standard values of $\alpha = 0.05$ and $\beta = 0.8$. Mean per capita consumption for final realised control group households is R3229.48 (lower than the phase 1 sample of R3491.037), with a standard deviation of R6461.36 (more than double that of the phase 1 control group). Two sample ranges are considered: a sample of 6 or 7 CSM households each for 500 claims ($n = 3000 - 3500$); and the rounded down sample achieved of 510 claims of 5 or 6 CSM households each ($n = 2550 - 3060$). The actual sample has 511 claims with a mean cluster size of 5.23 and 2670 households. We use 510 instead of 511 claims as only even numbered samples can be used to calibrate sample sizes for a clustered design. ρ denotes the intra-cluster correlation (ICC). The ICC is a measure of the extent to which households within any given claim can be expected to have outcomes that are similar. For instance $\rho \approx 1$ means that every household in a claim has virtually the same level of consumption. By contrast $\rho \approx 0$ means that there is virtually no correlation between the consumption levels of households in a claim. We have allowed for a wide range of ICC values, with an upper limit of 0.15. The realised ICC is 0.13 and is also shown in the table. Estimates represent minimum detectable effects (MDEs). The MDE is an estimate of the smallest level of impact that the applicable design is capable of detecting.

3.3.2 Psychological well being and cognition

Figures 3.5 as well as Table A.1 (Appendix A) show the estimated MDEs for the CES-D-10 score, expressed in standard deviation terms, using the same sample size calibrations as for consumption.

A similar pattern is evident in the gradients of these curves as compared with those for consumption: power increases steeply for a smaller number of sampled households per claim and then flatten out. Remarkably, for the targeted samples shown in the sensitivity analysis, the CES-D-10 measure shows very high expected precision. This is shown more clearly in Table A.1 where we see the design is powered to detect very small changes in depression scores (between 0.10 and 0.14 of a standard deviation), for a relatively large increase in the ICC from 0 to 0.2.

Very similar patterns are obtained for all of the cognitive outcomes. One example of those outcomes is the Corsi block test of working memory. Performance in this task is measured by the so-called "span" of the respondent. The span measures the number of objects the respondent has to hold in memory, which in our case is the order of blocks flashing at random on a screen. The results from Phase 1 of the data collection indicates that on average, respondents could correctly recall the sequencing of the flashing blocks if on average 4.6 blocks were involved in the sequence. These averages are far too small to use in MDE calculations that are based on the normal distribution as the empirical distributions do not satisfy that assumption. Instead, we convert the span score into a proportion and use the binomial distribution instead of the normal distribution. The results are given in Appendix A, Table A.2 and Figure A.1. The realised sample is capable of detecting 6-8% increases in working memory scores.

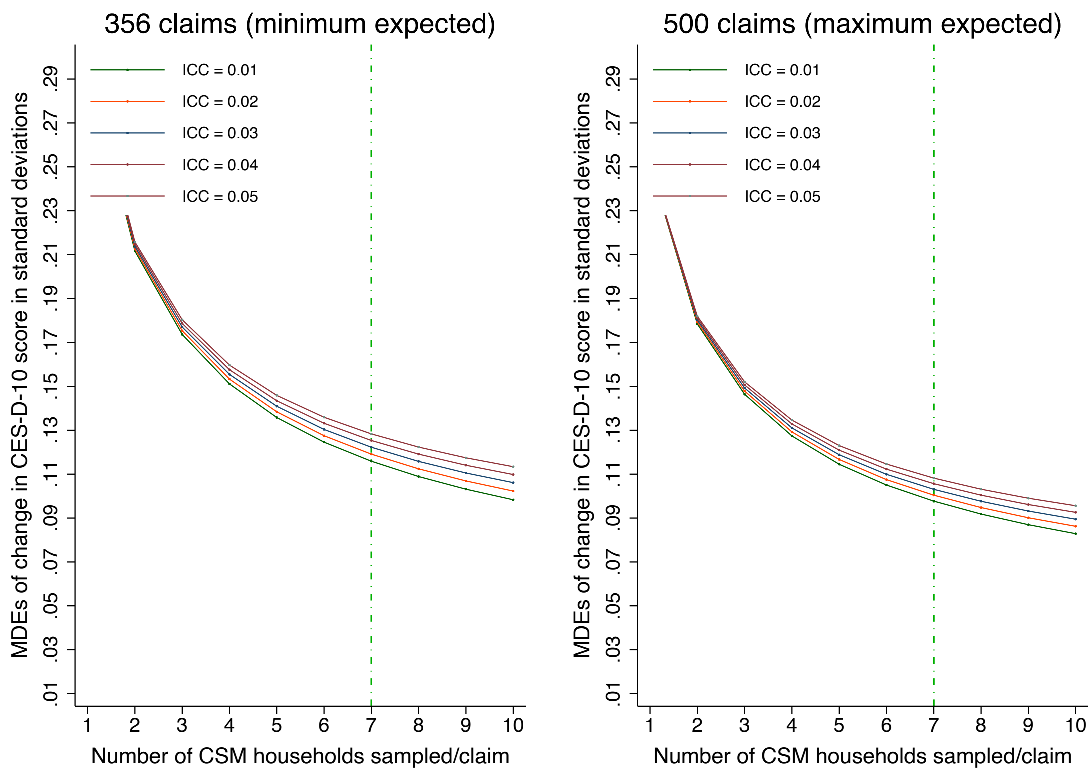


Figure 3.5: Minimum detectable effects for CES-D-10 Depression Score

Test size and power: we assume standard values of $\alpha = 0.05$ and $\beta = 0.8$. Mean depression score for control group households estimated from Phase 1 of data collection is 9.04, with a standard deviation of 5.51. Two sample ranges are considered: a smaller sample of $1-10$ CSM households each for 356 claims ($n = 356-3560$); and a larger sample of $1-10$ CSM households each for 500 claims ($n = 500-5000$). ICC denotes the intra-cluster correlation. Estimates represents minimum detectable effects (MDEs).

3.4 Statistical methodology

The study design is quasi-experimental.[7] Quasi-experimental methodologies retain the feature of a treatment-control comparison. However the allocation to each group is not conducted through a simple randomisation process, but is pre-determined. If the assignment to each group is not randomised, then before the treatment effect (impact) can be estimated, we have to correct for the bias caused by non-random assignment. This bias is called "selection bias." When working with quasi-experimental data, we have to decide how to simulate what would have been different had we followed a randomisation approach. There are many statistical and econometric approaches to solving this problem. We use a technique called generalised propensity score conditioning [46–48]. In this section, we present a non-technical description of how this technique works. The full details of the approach are given in an accompanying technical appendix to the report (Appendix B).

[7]The original design of LRES followed a pipeline randomisation methodology. However, implementing this design within the timeframe of the study became impossible due to a two year shutdown of fieldwork during Covid19 lockdown restrictions. During this time, a no-cost extension of the study was negotiated and agreed to with the Commission under a new quasi-experimental design. This agreement was also noted by the study Steering Committee. The operations close-out report to the Commission details this process more fully.

This technique is based on the idea of the propensity score: i.e., the probability (propensity) to be in the treatment group [49]. In the basic set-up, the treatment is binary: people either are in the treatment group, or they are in the control group. The task then is to estimate the probability of being in the treatment group and then to use the estimated propensity scores to correct for selection bias. There are many ways to implement this approach. A common approach is to match treated and control individuals according to their propensity scores. For example, if two individuals have approximately the same probability of being in the treatment group, but only one individual is actually observed to be in the treatment group, the one that is in the control group is considered a good comparison. Using this principle, every individual that is observed to be treated can be matched with one or more individuals who is not treated but that has approximately the same probability of being treated, but was not actually treated. This type of approach is called propensity score matching (see Appendix B.1 for the complete technical details).

The approach we use is based on the idea of the propensity score, but differs from the basic method described above in two important ways. The first difference is that the treatment we mostly focus on is not binary, but continuous. Instead of estimating average treatment effects for the treated, we estimate average treatment effects for different *levels* of treatment. This means that we do not simply compare outcomes between those who have received restitution (treatment group), against those awaiting restitution (control group), but we take the comparison a step further by comparing outcomes among those who receive different levels of restitution, as measured by the value of the restitution award that was made. Therefore, the propensity score in our implementation of the approach is not a probability, but rather an estimate of what the award amount could have been, controlling for the characteristics of the individual and important factors about the context of the claim to which the individual belongs.

The second difference is that we do not match individuals based on their propensity scores. Instead, the process is more straightforward. We first run a linear regression model of the outcome of interest, conditional on the propensity score and restituted award value. Then using the estimated regression coefficients from this model, we calculate the mean outcome for different award amounts represented in the data. The treatment effect is then shown by plotting these conditional means against various points of the distribution of award values (see Appendix B.2 for the complete technical details).

Section 4

Data

In this section of the report, we present summary statistics of the data. LRES collected a trove of administrative, survey, and geo-spatial data. However, for the purposes of this report we only discuss variables that appear directly in the models that we used to estimate impacts.[1] This is partly to ensure that we keep the focus on the modelling choices we made so that the treatment effects are clearly presented. We begin by describing the relevant variables in the models, and the rationale for choosing them. This is covered in Section 4.1 below. We then show in Section 4.2 that these variables are key predictors of treatment and therefore need to be dealt with appropriately when estimating treatment effects.

4.1 Descriptive statistics

The full sample of data comprises 2664 households (1054 in the control group and 1610 in the treatment group); 3378 individuals (1757 in the control group and 1621 in the treatment group), spread across 505 claims (150 claims in the control group and 355 claims in the treatment group).

Table 4.1 shows the means and standard deviations (in parentheses) of the key household level variables used to estimate the generalised propensity score models of the next section. Appendix Table A.3 provides the individual level descriptive statistics. The first 10 rows of Table 4.1 shows the proportions of awards in the study sample for the past 10 years. These variables, of course, are only defined for for the treatment group. Approximately 74% of the awards in the sample were made at least 4 years ago (for instance 17.6% and 22% for 2017 and 2018 respectively). The table also shows that household size is more or less the same between the treatment and control group (approximately 4.8-4.9 individuals per household).[2] In terms of demographics, about 70% of the sampled households are African, 24% are Coloured, 3% are Indian, and about 2% are White.

Turning now to variables hypothesised to play a role in predicting treatment, Table 4.1 reports whether the household has at least one member that speaks English (42%); at least one member that is employed (56%); and at least one member that reports access to government (25%), a political party (26%) or legal advice (46%).[3]

[1]The metadata of the study, once published, will contain a full listing of all the data collected, valid ranges of variables, and code listings. Our focus on this report is only the the three impact domains as per the agreed change of scope to the study

[2]We have not show tests of differences in means at this stage, as the treatment effects were have focused on are not simply binary ones. We show the appropriate test we use in the next section.

[3]Access to government and political parties is defined as whether the household has at least one member who themselves are members of government or a political party, or whether they know someone who is. Access to legal advice is defined as

Table 4.1: Summary statistics: household level

	0	1	Total
Award made in 2013	.	0.0338	0.0338
	(.)	(0.181)	(0.181)
Award made in 2014	.	0.119	0.119
	(.)	(0.324)	(0.324)
Award made in 2015	.	0.0632	0.0632
	(.)	(0.243)	(0.243)
Award made in 2016	.	0.132	0.132
	(.)	(0.339)	(0.339)
Award made in 2017	.	0.176	0.176
	(.)	(0.381)	(0.381)
Award made in 2018	.	0.220	0.220
	(.)	(0.414)	(0.414)
Award made in 2019	.	0.0889	0.0889
	(.)	(0.285)	(0.285)
Award made in 2020	.	0.0757	0.0757
	(.)	(0.265)	(0.265)
Award made in 2021	.	0.0594	0.0594
	(.)	(0.237)	(0.237)
Award made in 2022	.	0.0325	0.0325
	(.)	(0.177)	(0.177)
Household size	4.888	4.839	4.858
	(3.429)	(2.648)	(2.981)
Opposition control in ward of household: 2011-2015	0	0.109	0.0658
	(0)	(0.312)	(0.248)
Opposition control in ward of household: 2016-2020	0	0.342	0.206
	(0)	(0.474)	(0.405)
Opposition control in ward of dispossessed property: 2011-2015	0	0.180	0.108
	(0)	(0.384)	(0.311)
Opposition control in ward of dispossessed property: 2016-2020	0	0.544	0.328
	(0)	(0.498)	(0.470)
Opposition control in ward of dispossessed property: 2021-	0.214	0.0607	0.121
	(0.410)	(0.239)	(0.327)
Race (1= African)	0.784	0.643	0.698
	(0.411)	(0.479)	(0.459)
Race (1= Coloured)	0.195	0.274	0.243
	(0.394)	(0.446)	(0.428)
Race (1= Indian)	0.0150	0.0469	0.0344
	(0.120)	(0.211)	(0.182)
At least 1 English speaker in household	0.223	0.547	0.419
	(0.417)	(0.498)	(0.493)
At least 1 employed person in household	0.674	0.492	0.564
	(0.469)	(0.500)	(0.496)
Access to government	0.170	0.307	0.253
	(0.376)	(0.462)	(0.435)
Access to a political party	0.176	0.320	0.263
	(0.381)	(0.467)	(0.440)
Access to legal advice	0.259	0.598	0.463
	(0.438)	(0.491)	(0.499)
Real per capita consumption in Rands (x1000)	3.229	4.123	3.769
	(6.461)	(7.000)	(6.804)
Real value of restitution award in Rands (x1000)	.	37.54	37.54
	(.)	(72.63)	(72.63)
Observations	2646		

Access to government and political parties is defined as whether the household has at least one member who themselves are members of government or a political party, or whether they know someone who is. *Access to legal advice* is defined as whether the household has at least one person capable of providing legal advice or whether they know someone who can. The variable *Opposition control in ward of dispossessed property in 2011* is a dummy variable that measures whether the claim was finalised in the years between the 2011 and 2016 municipal elections and the dispossessed property was in a ward where opposition parties won the majority of seats in 2011. Similarly the 2016, and 2021 versions of this variable indicate claims finalised between 2016 and 2020, and 2021-2022 respectively and where opposition parties won the majority of seats in 2016 and 2021.

The final block of variables shown in Table 4.1 are meant to capture the effect of the local political environment. The political environment is often strongly related to the scale and scope of land reforms [50]. To proxy for the local political environment, we use public access data on the last three municipal elections; 2011, 2016 and 2021. We construct two categories of polity variables. The first category measures opposition control in the ward of the beneficiary household. The second measures whether the property being claimed as a result of historical dispossession is in a ward where opposition political parties have control. The variable *Opposition control in ward of dispossessed property in 2011* for instance, is a dummy variable that measures whether the claim was finalised in the years between the 2011 and 2016 municipal elections and the dispossessed property was in a ward where opposition parties won the majority of seats in 2011. Similarly the 2016 version of this variable is a dummy variable for claims finalised between 2016 and 2020 and where opposition parties won the majority of seats in 2016. Table 4.1 shows, for instance, that for the period 2016-2020, more than half (54%) of the dispossessed properties applicable to the sample of treated households were located in wards where opposition parties were in control.

We hypothesise that the dispossessed land polity variables are key determinants of award size for several reasons. First, it is often the case that dispossessed properties that are located in wards under opposition control tend to be harder to restitute, as these wards are generally wealthier, which means the dispossessed property is therefore likely to be of much higher value thus making it harder for the state to purchase the land. In this type of situation, the state normally searches for an alternative property of similar value. This often does not lead to a solution. The recourse then is to try to settle the claim through financial compensation. Historically, this type of situation leads the claims to be very small in value, as they are based on a historical valuation of the property rather than its current market value. This tends to be the rule rather than the exception. To drive home the logic, we present a counter example to the rule, in the box below: the case of Protea Village in the affluent suburb of Bishopscourt in Cape Town. The example of Protea Village is a rare instance where a high value dispossessed property in a ward under opposition control was eventually transferred back to the beneficiaries, so we present this case in a box.

whether the household has at least one person capable of providing legal advice or whether they know someone who can.

The case of Protea Village

The community of Protea Village lodged a claim for restitution of their dispossessed property that is located in the highly affluent suburb of Bishopscourt in the city of Cape Town, where the official opposition party is in control. This claim was originally settled through cash compensation about 20 years ago. However, 86 families refused the cash awards. This led to protracted legal battles, aided by the state, that saw their rights to the actual land restored. However the residents association of Bishopscourt challenged the claim. The court decision found in favour of the Protea Village Community in 2022 and the planning process to build houses on the land is only now beginning.

The location of the Protea Village property surely does plays a role in determining the outcome of the claim. It would be no coincidence to learn that the attitudes of the wealthy residents that tried to block the claim probably reflect those of the opposition party to some degree. Since electoral outcomes, at least in South Africa, don't directly affect beneficiary outcomes, the presumption that polity variables are exogenous seems reasonable. But we would not expect such variables to be very strong predictors in the case of Protea Village because the vast majority of the families that accepted the cash compensation received far less compared to the value of the restored rights to the 86 families that held out. These two opposing effects potentially cancel out any correlation we might expect between opposition control and award size.

To better understand the unique example of Protea Village, we undertook a detailed case study of the group. We interviewed representatives of 63 of the 86 families to better understand how it differed from other communities that did not prevail in quite the same way. We mapped their social connectedness with one another and measured their civic attitudes. As detailed in our companion case study report, many interesting features of this community – how they cooperate, how they share information etc., – sets it apart from the other case studies we conducted. There is clearly something unique about this community that shaped its resolve and ultimate success at securing their rights.

But this community is the exception. More often, we expect the opposite. Ordinarily, we would expect that more powerful, well resourced neighbourhoods that fall under opposition control will exert a limiting effect on the scope and scale of restitution, thus increasing the likelihood that settlements take the form of cash compensations based on the historical valuation of the land, rather than it's current market value. This key intuition guides our approach to the use of polity variables. These variables are expected to be negatively related to award size and are plausibly exogenous to the outcomes we wish to measure. This last requirement is key to using propensity score methods to purge selection bias from impact estimates.

4.2 Covariate balance

Land reforms ultimately stem from political processes, whether in an effort to rebuild society after civil conflict, or as the direct result of political transformation [51, 52]. The changes in legislation that mandate the scope of land reforms is often a useful place to start when thinking about how to measure the impacts of the targeted land reform interventions [53, 54]. Although restitution programmes contribute a small fraction to land reforms world wide, where such programmes have been undertaken and carefully studied, the local political landscape turns out to be a key driver [55]. The scope (or demand) for land reforms can therefore be modelled as a function of local politics [50]. We use this key insight to build our rationale for the choice of covariates used to estimate the the generalised propensity score (GPS), and then show that controlling for the GPS balances the data.

4.2.1 Generalised propensity scores (GPS)

Recall that in order to estimate an impact from quasi-experimental data, we cannot assume there is no selection bias present in the data. The so-called "naive" estimate of the treatment effect; the straightforward difference in mean outcomes between the treatment and control group, would give a misleading picture of impacts. The potential for selection bias must be tackled through the choice of appropriate statistical methodologies. As explained in section 3, our choice of technique is the generalised propensity score approach.

Tables 4.2 and 4.3 show the results of estimating the generalised propensity score model for household level variables and individual level variables. The tables report only the key covariates that we hypothesise to be important determinants of award size. The full set of results is shown in Appendix A, Tables A.4–A.5.[4]

Model 1 of Table 4.2 controls only for household characteristics that we hypothesise to be relevant to award size. The first block of covariates are binary indicators for the race group to which the household belongs. [5]The excluded race category is white.[6] We also control for whether the household contains at least one person that speaks english as a home language. There is obviously a strong overlap between being a home language english speaker and race. Thus we do not find it surprising that the coefficient for *Race (1 = African)* is not statistically significant. By contrast, Coloured and Indian beneficiaries receive smaller and larger awards, respectively, relative to White beneficiaries. What is somewhat more surprising is that being an home language English speaker is negatively related to award size. This is probably due to historical accident; some of the most high value claims over the last 10 years have been made to African beneficiaries, and we know from the last two population censuses that speaking English as a home language is higher among both Indians and Coloureds than among Africans.

A puzzling finding is the negative relationship between employment in the beneficiary household and award size. We do not wish to over interpret this finding. It is possible that a similar logic to the english language variable applies here. The final set of controls relate to forms of agency that the beneficiary has at their disposal that could facilitate higher awards. The variable *Access to government* is positive and

[4]The key difference between the two sets of tables is that the coefficients for year of award are only shown in the versions reported in the appendix. Most of these coefficients are insignificant except for 2019 and 2020 which are significant at the 10% level. This finding suggests that the values of awards over the 10 year period covered by the data (2013-2022) do not changed notably, except in 2019 and 2020.

[5]In the sample of data we collected, there are no mixed race households. This need not be true for the population of restitution beneficiaries in general, but given that the sample of claims we collected data from is large relative to the number of outstanding claims

[6]While the vast majority of restitution beneficiaries are non-white families, there is a small number of white households in the sample (64 in total). This is completely possible as the restitution legislation does not preclude white households from claiming compensations if they meet the vetting criteria laid out in Appendix C.4.

Table 4.2: Maximum likelihood estimates of the determinants of award size (household level)

Variables	Model 1	Model 2
Race (1= African)	0.0313 (0.16)	-0.163 (-0.81)
Race (1= Coloured)	-0.520** (-2.68)	-0.584** (-3.05)
Race (1= Indian)	0.740** (3.04)	0.634** (2.64)
At least 1 English speaker in household	-0.371*** (-4.72)	-0.352*** (-4.57)
At least 1 employed person in household	-0.226** (-3.22)	-0.242*** (-3.52)
Access to government	0.167 (1.91)	0.149 (1.75)
Access to a political party	-0.120 (-1.38)	-0.135 (-1.59)
Access to legal advice	-0.00875 (-0.12)	0.00155 (0.02)
Household size		0.00174 (0.13)
Opposition control in ward of household: 2011		-0.305 (-1.92)
Opposition control in ward of household: 2016		-0.0351 (-0.35)
Opposition control in ward of dispossessed property: 2011		-0.338 (-1.67)
Opposition control in ward of dispossessed property: 2016		-0.225* (-2.17)
Opposition control in ward of dispossessed property: 2021		0.244 (0.85)
Constant	3.012*** (15.00)	3.470*** (11.09)
Observations	1574	1564

Access to government and political parties is defined as whether the household has at least one member who themselves are members of government or a political party, or whether they know someone who is. *Access to legal advice* is defined as whether the household has at least one person capable of providing legal advice or whether they know someone who can. The variable *Opposition control in ward of dispossessed property in 2011* is a dummy variable that measures whether the claim was finalised in the years between the 2011 and 2016 municipal elections and the dispossessed property was in a ward where opposition parties won the majority of seats in 2011. Similarly the 2016, and 2021 versions of this variable indicate claims finalised between 2016 and 2020, and 2021-2022 respectively and where opposition parties won the majority of seats in 2016 and 2021.

statistically significant (but only at the 10% level). Surprisingly, access to legal advice does not appear to be correlated with the size of award.

In Model 2 we introduce the polity variables. In Table 4.3, which estimates similar models of award sizes to individuals rather than their households, these variables are included in both models. As we argued above, the political environment is often strongly related to the scale and scope of land reforms [50].

Both Tables 4.2 and 4.3 report negative and statistically significant effects of polity on award size. In both 2011 and 2016, opposition control both in the ward/neighbourhood of the beneficiary household as well as at the site of the claimed property cause reductions in award sizes compared as compared to beneficiary and claimed property locations that are not under opposition control.

4.2.2 Covariate balance conditional on the GPS

The variation in award sizes made to restitution beneficiaries can be thought of as primarily a random process, in the sense that two people with similar characteristics end up receiving different awards because of the way in which the laws of succession apply to any given beneficiary. Specifically, when the claim was lodged 25 or more years ago, a set of beneficiaries would have been identified. The time elapsed between lodgement and finalisation of a claim (a minimum of 15 years in the sampling frame) often leads to new beneficiaries being included in the claim because of the death of a parent who would have been the original beneficiary. That original right to restitution then passes to the deceased person's surviving descendants, and so the award amount is divided between the surviving children. If not all the children have survived to finalisation, that child's right then cedes to *their* surviving children (or some of the grand children of the person with the original right). This part of the variation in award size should, in principle, be uncorrelated to outcomes. In reality, it is not entirely random, because outcomes are correlated with age, and older beneficiaries get larger awards. However, our sampling strategy mitigates this by not sampling very old and very young beneficiaries. However, we would also want to partial out the age effect as well as any other factor likely to affect the restitution award amount. This is the primary role of the GPS.

The main idea is to use the GPS to show that once it is controlled for in estimating the conditional mean of the confounding variable, the confounding effect is removed. In other words, conditioning on the GPS balances the data. This is done by using the coefficients reported in Table 4.2 to estimate the GPS variable (i.e., the predicted values of award amounts). Then the conditional mean of each covariate is estimated, controlling for the GPS. Finally, the balance test is performed.

The implementation of the balance test we chose is based on terciles of the award distribution. The test is based on a test statistic (a Bayes factor) that tests that the covariate means (conditional on the GPS) is not different between units who belong to a particular treatment tercile interval and units who belong to all other treatment tercile intervals. We conducted this test for Model 2 for Table 4.2. Using Jefferys' order of magnitude criterion, a Bayes factor of greater than 1 is decisive evidence that conditional on the generalised propensity score, the covariates are no longer correlated with the treatment (i.e., the GPS expunges the confound). A similar test is performed at the individual level using the results of Table 4.3. The results of both tests are reported in Appendix Tables A.6 and A.7. Both sets of results show that conditioning on the GPS decisively balances the cofounding effects out as all Bayes factors exceed 1.

Table 4.3: Maximum likelihood estimates of the determinants of award size (individual level)

Variables	Model 1	Model 2
Household size	0.00187 (0.14)	-0.00238 (-0.18)
Opposition control in ward of household: 2011	-0.303 (-1.92)	-0.279 (-1.77)
Opposition control in ward of household: 2016	-0.0357 (-0.36)	-0.0419 (-0.42)
Opposition control in ward of dispossessed property: 2011	-0.343 (-1.71)	-0.396* (-1.98)
Opposition control in ward of dispossessed property: 2016	-0.226* (-2.18)	-0.222* (-2.16)
Opposition control in ward of dispossessed property: 2021	0.240 (0.84)	0.239 (0.84)
Race (1= African)	-0.168 (-0.84)	-0.189 (-0.95)
Race (1= Coloured)	-0.604** (-3.17)	-0.626*** (-3.30)
Race (1= Indian)	0.630** (2.63)	0.592* (2.48)
At least 1 English speaker in household	-0.356*** (-4.65)	-0.310*** (-3.91)
At least 1 employed person in household	-0.245*** (-3.58)	-0.229** (-3.29)
Access to government	0.136 (1.61)	0.148 (1.76)
Access to a political party	-0.133 (-1.57)	-0.132 (-1.58)
Access to legal advice	-0.000690 (-0.01)	0.00681 (0.09)
Gender is male		0.178** (2.59)
Education (years)		-0.0324** (-3.04)
Observations	1575	1575

Access to government and political parties is defined as whether the household has at least one member who themselves are members of government or a political party, or whether they know someone who is. Access to legal advice is defined as whether the household has at least one person capable of providing legal advice or whether they know someone who can. The variable *Opposition control in ward of dispossessed property in 2011* is a dummy variable that measures whether the claim was finalised in the years between the 2011 and 2016 municipal elections and the dispossessed property was in a ward where opposition parties won the majority of seats in 2011. Similarly the 2016, and 2021 versions of this variable indicate claims finalised between 2016 and 2020, and 2021-2022 respectively and where opposition parties won the majority of seats in 2016 and 2021.

Section 5

Results

5.1 Impacts on consumption

The raw mean of per-capita consumption is R894 higher in the treatment group compared to the control group. This constitutes a naive treatment effect of about 28%. This difference is highly statistically significant (the MDE for the realised sample is 23.54%). Whiles these estimates are useful benchmarks in terms of assessing the power of the study, they are limited in that we would expect outcomes to differ by award amount. To address this issue, we estimate treatment effects *for different amounts of restitution awards*; i.e., the treatment becomes a continuous variable.

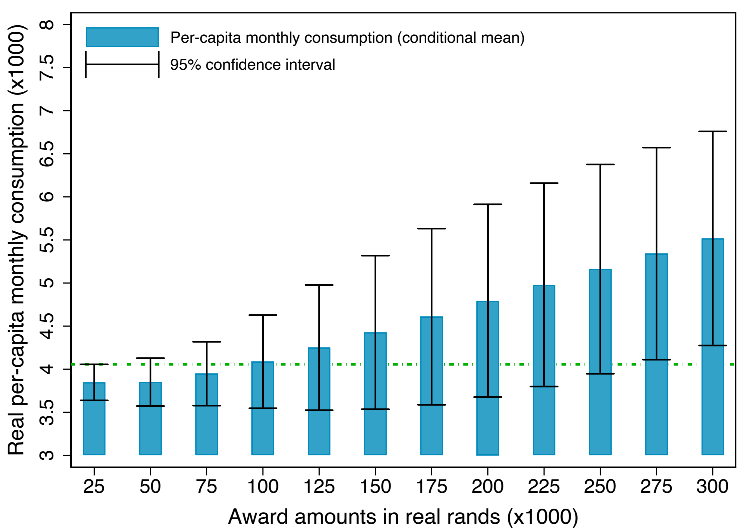


Figure 5.1: Treatment effects: per-capita consumption (only beneficiary controls)

The graph is obtained by estimating a linear regression (quadratic specification) of per-capita consumption, Y, controlling for award value, d, the generalised propensity score r, and the interaction between d and r. The coefficients from this regression are then used to estimate $\widehat{\psi}(d) = E[\widehat{Y(d)}] = \frac{1}{n_1}\sum_{i=1}^{n_1}(\widehat{\alpha}_0 + \widehat{\alpha}_1 \cdot d + \widehat{\alpha}_2 \cdot d^2 + \widehat{\alpha}_3 \cdot \widehat{r}(d, \mathbf{x}_i) + \widehat{\alpha}_4 \cdot \widehat{r}(d, \mathbf{x}_i)^2 + \widehat{\alpha}_5 \cdot d \cdot \widehat{r}(d, \mathbf{x}_i))$ for 12 different values of d, ranging between 25000 and 300000. Standard errors are bootstrapped. These are then used to compute the 95% confidence bands.

Figures 5.1 and 5.2 show the main treatment effects for the two specifications of the generalised propensity models reported in Table 4.2. Both graphs represent graphical versions of the treatment effect model of equation 5.8. It shows the estimated averages for the main outcome variable (per capita consumption, measured in units of R1000) for 14 different award amounts, ranging from R25000 per household to R300000 per household. For example, Figure 5.1 shows that per capita consumption is about R4000 for an award amount of R100000, and rises to a maximum of about R5500 for an award amount of R300000.[1]

In the binary treatment case, the treatment effect is the difference in the outcome variable between the treatment group and the control group. However, when we instead choose to focus on impacts of different award amounts, the counterfactual is not the average outcome for the control group, but rather the average outcome for those receiving a specified benchmark level of treatment. We choose this benchmark treatment level to be R25000. For this award amount, the bias-adjusted mean of per-capita consumption is R3887 and the upper-limit of the 95% confidence interval is R4024. The horizontal reference line in both Figures 5.1 and 5.2 is this upper limit. A valid treatment effect (as opposed to a null finding) are award amounts where the bias-adjusted lower limit of the 95% confidence interval of the effect does not overlap with the upper limit estimate for an award of R25000.

Figure 5.1 is based on a limited set of controls for selection bias (Model 1 of Table 4.2). Model 2 of Table 4.2 however, considers the larger set of control variables that include the polity variables. The latter is better performing at stripping out other selection biases that are unrelated to beneficiary characteristics). As already discussed, proxies for the local political environment are plausible determinants of restitution demand. Using the generalised propensity score variable of this model, we reestimate the treatment effects on the treated (Figure 5.2). The control variables underlying these estimates are perfectly balanced, as discussed earlier (all have Bayes factors that exceed their threshold levels of significance).

The estimated impacts are much larger for the expanded model of the GPS, albeit with much larger confidence bands for very large award amounts. Two findings are apparent. First, the award response function is upward sloping for all levels of awards considered. Second, significant treatment effects are estimated for award amounts of R200000 and higher. For instance, the impact on per capital consumption, of receiving R300000 in restitution compensation is approximately 48%, relative to an award amount of R25000.[2] This finding is surprising only in magnitude. As mentioned previously, the raw mean for the treatment group is R894 higher than that of the control group. This difference is highly statistically significant (t statistic of 9). This represents approximately 31% of the mean per capita real consumption for the larger South African population as measured over fives waves of the National Income Dynamics Study.[3]

The implied impacts of the standard settlement amount relative to an award amount of R25000 is higher but not statistically significant. However the Commission recently increased the standard settlement offer to approximately R200000.[4] From a policy standpoint, it makes sense to take this amount as the appropriate treatment level. For this award amount, compared to an award amount of R25000, we estimate a treatment effect of 25-26%. This is quite a large and statistically significant effect. However when compared against the raw mean difference of 31% between the treatment and control group, the finding is not altogether surprising.

[1] We do not estimate treatment effects for award amounts above R300000, which is at the 98th percentile of the award distribution. Figure A.2 shows the distribution of awards up to R300000.

[2] The bias-adjusted conditional mean for an award of R300000 is R5768 versus R3887 for an award of R25000.

[3] The average across all five waves is R2869.44 inflated to December 2022 prices.

[4] Some sources indicate that the new amount is R202000. Other sources indicate a slightly higher amount of R208000.

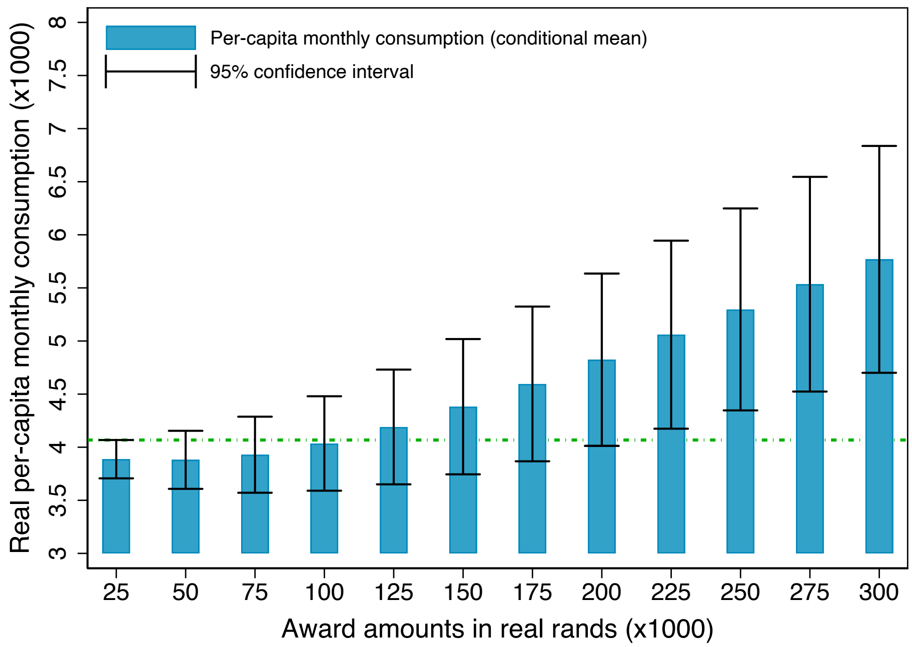


Figure 5.2: Treatment effects: per-capita consumption (full set of controls)

The graph is obtained by estimating a linear regression (quadratic specification) of per-capita consumption, Y, controlling for award value, d, the generalised propensity score r, and the interaction between d and r. The coefficients from this regression are then used to estimate $\widehat{\psi}(d) = \widehat{E[Y(d)]} = \frac{1}{n_1}\sum_{i=1}^{n_1}(\widehat{\alpha}_0 + \widehat{\alpha}_1 \cdot d + \widehat{\alpha}_2 \cdot d^2 + \widehat{\alpha}_3 \cdot \widehat{r}(d, \mathbf{x}_i) + \widehat{\alpha}_4 \cdot \widehat{r}(d, \mathbf{x}_i)^2 + \widehat{\alpha}_5 \cdot d \cdot \widehat{r}(d, \mathbf{x}_i))$ for 12 different values of d, ranging between 25000 and 300000. Standard errors are bootstrapped. These are then used to compute the 95% confidence bands.

5.2 Impacts on psychological well-being

We do not present binary treatment results for this outcome because the depression scores for the treatment group (mean = 8.63; s.d = 5.42; Table A.3) are very close to that of the control group (mean = 8.53; s.d = 6.10; A.3). This difference in standard deviations is less than - 0.02. Moreover, the cut-off for the CES-D-10 screen is 10 [41, 56]. This implies that neither the treatment group nor the control group are at risk of depression on average.

This is an interesting and somewhat surprising finding that runs contrary to our hypothesis that the programme should have improved psychological well being measurably. To put these estimates into perspective, the average CES-D-10 score across 5 waves of data collected by the National Income Dynamics Study is 7.16. So even though the population of restitution beneficiaries do not appear at risk of depression, they are indeed at higher risk than the general population. A further explanation for the very small difference between the treatment and control groups relates to the stage at which we have measured their depression risks. For both groups, the long process of awaiting their restitution awards was at an end. This fact of the design makes it possible for the psychological benefits afforded by the finalisation of the claim to have been absorbed prior to when they completed the depression screen. Since a major component of affective states is anticipatory, we cannot rule out this possibility. Both depression and anxiety can be thought of in this way [57]. Worries, when held for a long time, can become entrenched and could lead to depression and/or anxiety. If a strong determinant of restitution claimants' affective states are linked to the long wait to see their claims approved, the act of settlement itself could cause the affective states of the control group to

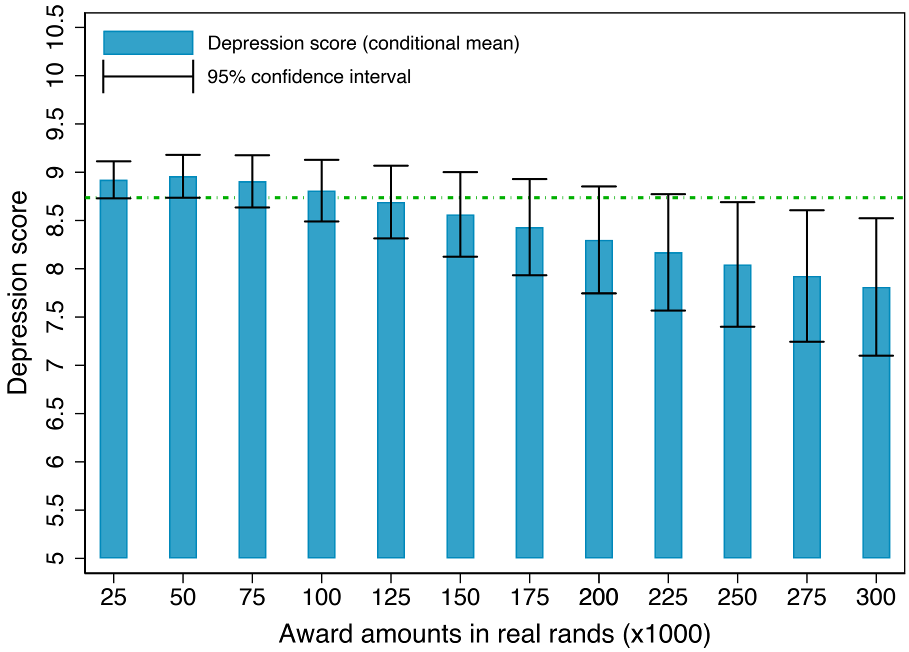


Figure 5.3: Treatment effects: CES-D-10 depression score

The graph is obtained by estimating a linear regression (quadratic specification) of the CES-D-10 depression scoe, Y, controlling for award value, d, the generalised propensity score r, and the interaction between d and r. The coefficients from this regression are then used to estimate $\widehat{\psi}(d) = \widehat{E[Y(d)]} = \frac{1}{n_1}\sum_{i=1}^{n_1}(\widehat{\alpha}_0 + \widehat{\alpha}_1 \cdot d + \widehat{\alpha}_2 \cdot d^2 + \widehat{\alpha}_3 \cdot \widehat{r}(d, \mathbf{x}_i) + \widehat{\alpha}_4 \cdot \widehat{r}(d, \mathbf{x}_i)^2 + \widehat{\alpha}_5 \cdot d \cdot \widehat{r}(d, \mathbf{x}_i))$ for 12 different values of d, ranging between 25000 and 300000. Standard errors are bootstrapped. These are then used to compute the 95% confidence bands.

shift closer to the treatment group. Since the sampling for the study screened out claims that were not yet settled, we can not test for anticipatory effects for depression.

However, our empirical strategy does allow us to ask and answer the bigger question of whether psychological well being improves in the long run. Consistent with our approach to estimating impacts for household well-being, the impacts for different levels of award sizes speaks to this question, because we collected data for 10 years of restitution payments spanning 2013-2022. The GPS model controls for the year of receipt of the award, so the treatment effects we estimate conditional on the GPS can be interpreted as a long run (i.e., time-invariant) effect. Figure 5.3 shows the main results (see Table 4.3 for a summary of the GPS model and Appendix, A Table A.5 for the full model). As before, the GPS balances the data; all Bayes factors exceed the threshold of 1 (Appendix Table A.7). The results show that the award response function is negative, implying that increases in award amounts cause a reduction in depression scores.

The results are mixed; we can't say much about impacts for very small award sizes as the estimates are not very precise for small award amounts, but for the larger award values considered (R250000 and higher), there is clear cut evidence of strong impacts. The bias-adjusted depression score for an award of R250000 is 8.96 compared to 8.04 for an award of R50000. This almost 1 point decrease in depression scores is about 0.15 of a standard deviation and represents about a 10% reduction in the risk for depression.

5.3 Impacts on cognition

In this final section, we look at the findings for the working memory task.[5] Figure 5.4 shows the main findings.

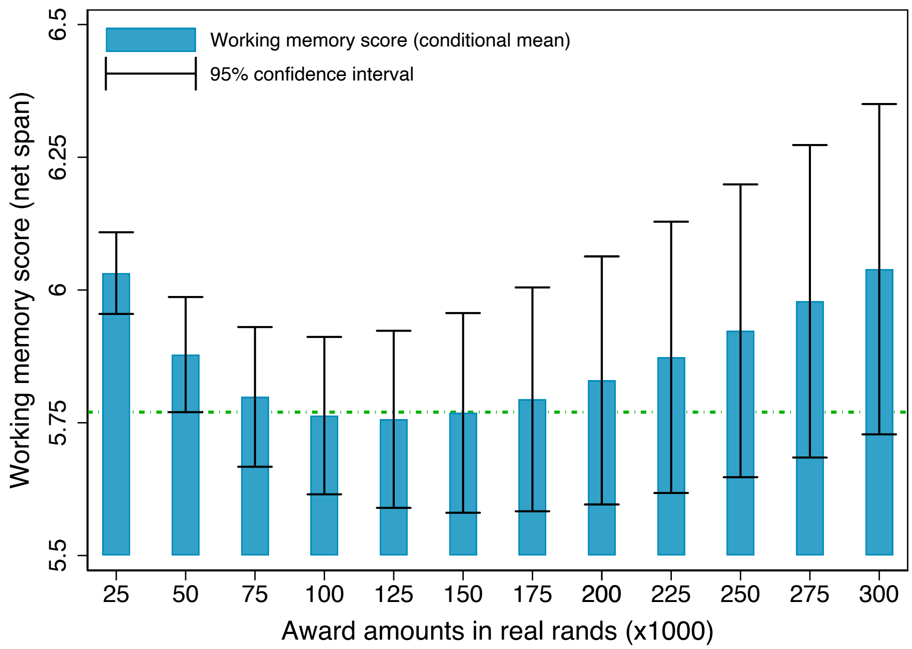


Figure 5.4: Treatment effects: working memory score

The graph is obtained by estimating a linear regression (quadratic specification) of working memory span, Y, controlling for award value, d, the generalised propensity score r, and the interaction between d and r. The coefficients from this regression are then used to estimate $\widehat{\psi}(d) = \widehat{E[Y(d)]} = \frac{1}{n_1}\sum_{i=1}^{n_1}(\widehat{\alpha}_0 + \widehat{\alpha}_1 \cdot d + \widehat{\alpha}_2 \cdot d^2 + \widehat{\alpha}_3 \cdot \widehat{r}(d, \mathbf{x}_i) + \widehat{\alpha}_4 \cdot \widehat{r}(d, \mathbf{x}_i)^2 + \widehat{\alpha}_5 \cdot d \cdot \widehat{r}(d, \mathbf{x}_i))$ for 12 different values of d, ranging between 25000 and 300000. Standard errors are bootstrapped. These are then used to compute the 95% confidence bands.

The naive binary treatment effect is 0.87 of a standard deviation (mean span is 5.21 for the control group and 6.08 for the treatment group; Table A.3). This naive estimate is highly statistically significant ($t-$ratio of 9.09, not shown). The continuous treatment estimates shown in Figure 5.4 are not significant; although there is a clear positive effect of higher awards on performance in the working memory task, the confidence bands are generally too large to distinguish these positive effects from null effects. There also appears to be a negative relationship for award amounts of less than R100000. These patterns will therefore need further exploration in future work.

[5] We deployed several other cognitive tests in Phase 1 as discussed in the chapter on measurement. However, only the working memory task could be deployed in Phase 2 due to the limitations imposed by moving to telephonic data collection. However we took the opportunity to reintroduce the full test battery in the Phase 2 qualitative case studies, which was conducted in person. This data has not yet been carefully analysed so we do not present those findings.

Section 6

Conclusion

6.1 The restitution-redress nexus

In her influential book, *Landmarked*, Sheryl Walker reflects on her experience as the Regional Land Claims Commissioner for KwaZulu-Natal during the first term of the Commission, 1995-2000. A central theme of her authoritative account is what she calls the problematic "master narrative" of loss and restoration. She writes:

> The simple story of forced removals leads to a narrative of restitution that is constructed around the equally ingenious idea of reversal. The task becomes simply to turn those elements of dispossession in the past around – to put ownership of land in the hands of 'the majority', restore those who were uprooted to the land from which they were torn, and reconcile those who lost with those who gained. Underpinning it is the naive hope that through this act of reversal our society will indeed reach its promised land and thereby overcome the entrenched poverty, suffering, alienation, ignorance and conflict we see all around us.

The master narrative, she notes, breaks down because it fails to consider both the broader contexts within with restitution takes place, and because it lacks the specificity required for defining success. That specificity, she argues, lies in a focus on the material conditions of beneficiaries lives as a main goal of restitution [29]. Her book offers three core case studies illustrating the role of social cohesion, both as a cause and a consequence of restitution, and the multi-dimensional ways in which beneficiary well-being can be affected, both positively and negatively. The book therefore represents an important shift in the framing of the case for restitution; away from national targets aimed at measuring the speed of processing claims, to one that is focused on beneficiary well being. Arguably, this is *the* main goal of any restitution programme.

6.2 Summary of main findings

The work undertaken by LRES and detailed in this report represents the first major, large scale, effort at taking up the challenge of adding specificity to measuring the impacts of the restitution programme. Empirical evidence of large magnitude transfers is quite scarce, with only a handful of studies to have emerged in the past decade. The restitution programme in South Africa is an ambitious programme when looked at from this vantage point. The average award size analysed in this evaluation is about R39000 or

5484 US Dollars PPP, which is more than 5 times the size of the transfers analysed in recent large magnitude cash or land transfer programmes worldwide [22–26].

In the preceding section, we reported striking evidence of a new narrative: specificity matters, even if we do not observe every element of what ought to be specified. We chose to ask simple and tractable questions that together gives us a detailed enough picture of beneficiary well-being, while at the same time ensuring the widest study sample ever collected on settled and finalised restitution claims; a total of 3735 individuals, across 2646 households from 505 claims, along with a further 6 claims for detailed case study.

We estimate that large magnitude transfers, whether in cash or land rights, causes sustained improvements to beneficiary well-being. Per capita consumption, the standard way to measure individual economic well-being, is estimated to increase by 25-26% in the long run, for award amounts of approximately R200000, compared to award amounts that are about 10 times smaller. The impact on psychological well-being, as measured by a decreased risk of depression, again for the same large and small award values, is estimated to be about 10% lower.

6.3 Study implications for restitution policy

6.3.1 Estimates can be interpreted as long-term impacts

With large magnitude transfers on this scale, it is virtually impossible to design an experiment that would allow identification of impacts, as one would need to both randomly assign different sizes of the transfer across beneficiaries at different points in time. The temporal aspect of any pure experimental design would be key, as short term impacts of large magnitude transfers are expected to affect welfare simply by dint of the income shock that the transfer would create. This is a key contribution of our evaluation of the South African land restitution programme which has key policy implications. By adopting a quasi experimental design, we are able to look back in time to awards made over the past ten years. By ensuring a good spread of awards over the 10 year period, we are able to rule out the possibility that our estimates are only detecting the short term income shock effect of the transfer. Since the underlying propensity score models we estimate all control for the year of award, the treatment effects we report can therefore be interpreted as long run impacts. This important point about interpretation should be deliberated by the Commission and other stakeholders in government as it could offer the basis for a further systematisation of how restitution awards are implemented.

6.3.2 An economic case for "equitable redress"

The impacts for large restitution awards are positive and sizable, by any measure. Further robustness testing will be undertaken as the findings are prepared for publication. This is a normally expected practice for impact evaluation studies and one should not be surprised if the impacts turn out smaller than reported here. There is also the possibility that they might be larger. While the current estimates are not surprising by themselves, they do challenge the priors and intuitions one might have about the longer term efficacy of the restitution programme. These are the first estimates we have of the (long-term) reach of the restitution programme based on a large sample of beneficiaries and claims. The Commission recently increased the standard settlement offer from R110000 to approximately R200000. A key policy takeaway, if not *the* key policy takeaway is that this decision makes good economic sense. For this award amount, compared to an award amount of R25000, the estimated impact on consumption is 25-26%. This is quite a large and statistically significant effect. If this finding survives further robustness testing, this would be a substantial level of impact and suggests an economic case for restitution. These findings for large award amounts (or equivalent value in restored land rights) affords stakeholders a platform to formalise an operational definition of "equitable redress". If historical valuations of dispossessed properties are generally unreliable, there might be scope to radically simplify this part of the process. The evidence suggests there are sizeable impacts of settlement awards that are more or less the same as the new standard settlement offer. This creates the space for operationalising the principle of "equitable redress" against this new evidence.

6.3.3 Innovating the use of small groups during options workshops

The lack of social cohesion among beneficiaries of a claim can be affected by restitution. But it can also be a strong impediment to the effective implementation of restitution. One such impediment that often turns out to be the key bottleneck in settling a claim concerns how beneficiaries go about choosing between the options of cash compensation versus restored land. The social (and quasi legal) structures through which this choice is made, are workshops run typically by the trustees of the a claim with the broader set of individual verified as beneficiaries. These workshops, and the various external stakeholders they often include (Commission staff, real estate developers and the various consultants they often employ) are staging grounds for how the "community" plans to come together. However, these groups are often very loosely defined communities, on the one end of the spectrum non-kin with no or little social connection to one another, and on the opposite end of the spectrum, members of one large extended family, who may or may not be close knit. The average claim will often lie somewhere in between these two extremes. The results from the six case studies we conducted, detailed in our companion report, has important implications for this component of the restitution process.

To examine the social cohesiveness of each of the six case study claims, we mapped the social network of each claim and identified distinct subgroups within each claim that share tight bonds with each other. Analysing the Protea Village network, we find high levels of cohesiveness and low levels of fragmentation (6 weakly defined subgroups). In contrast, Ndabeni showed very low social cohesion and high levels of fragmentation (20 strongly defined subgroups). These findings suggest very different community structures; in the case of Protea Village, a structure that would likely aid cooperation and conflict resolution; in Ndabeni the opposite.

As an exercise to test how this is likely to matter, we simulated the rate of information diffusion in each claim. The patterns of social cohesion in Protea Village claim predicts that if 10% of their beneficiaries were introduced to new information, within 5 steps of transmission, the information would on average reach between 35% and 40% of the members of this community. The same process in Ndabeni would only reach between 11% and 13%. The main message is that a community that resembles the type of social network structure of Ndabeni will find it hard to reach agreements. This comports well with the history of this claim which has experienced severe challenges since inception in 2004.

The main structural feature of the Ndabeni claim is its fragmentation. Knowing this type of information is key to promoting more effective, less drawn-out and therefore less conflict ridden engagements during the options workshops. Given that larger claims (like Ndabeni) will be more likely to exhibit more fragmentation, it makes sense to use the information about the network structure of claims to better execute how claims are settled. The main policy insight here is that by pre-identifying sub-groups within a claim, the Commission will be able to foresee and therefore mitigate hold-ups to settlement. One way is to disseminate information to each pre-identified subgroups and devise mixing strategies between subgroups during workshop engagements.

Using more granular data on the social ties between beneficiaries to better lubricate the discussions seems to be an easy tweak in implementation.

6.4 Study limitations

The discourse on restitution up to now, including the nuanced perspectives offered in Cheryl Walker's book cited earlier, offer narrow pieces of the story of the reach of restitution, as virtually all that discourse relies exclusively on small numbers of case studies. LRES is the first attempt to collect data from large samples of beneficiaries and claims so that it is now possible to form a more representative picture of how the lives of beneficiaries change in response to restitution.

The story from the systematised survey data collection is clearly a positive one; that of the qualitative case studies is more mixed, in line with previous work. In many ways, this is to be expected. To measure causal effects of the restitution programme (i.e., impacts) in tractable ways on beneficiary well-being, certain difficult choices had to be made, especially after the change in data collection modality from CAPI to CATI. There are therefore many other outcome measures that could have been investigated that have not been, because of the effects of COVID on in-person data collection. This limitation should be borne in mind.

The study also likely underestimates the impacts for households with multiple beneficiaries. To be able to gain traction on this problem, we would need to know how beneficiaries are related to one another at the point of drawing the sample. This requires building family trees for every sampled claim and developing procedures for grouping family members into co-residing household units. Since the treatment group could only be sampled from commitment registers and XXX lists, there would be no way to build family trees, as this is virtually impossible to do without first doing a census of each claim. This refinement to the analysis will have to be left to future research.

Bibliography

1. Platzky, L. & Walker, C. *The Surplus People: Forced Removals in South Africa* (Raven Press, Johannesburg, 1985).
2. Elisson, R. Harlem is Nowehere. *Harpers Magazine* **August** (1964).
3. Mayer, S. E. & Jencks, C. Growing up in poor neighborhoods: How much does it matter? *Science* **243,** 1441–1445 (1989).
4. Massey, D. S. American apartheid: Segregation and the making of the underclass. *American journal of sociology* **96,** 329–357 (1990).
5. Sampson, R. J., Raudenbush, S. W. & Earls, F. Neighborhoods and violent crime: A multilevel study of collective efficacy. *science* **277,** 918–924 (1997).
6. Sharkey, P. The acute effect of local homicides on children's cognitive performance. *Proceedings of the National Academy of Sciences* **107,** 11733–11738 (2010).
7. Ludwig, J. *et al.* Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults. *Science* **337,** 1505–1510. ISSN: 0036-8075. eprint: http://science.sciencemag.org/content/337/6101/1505.full.pdf. http://science.sciencemag.org/content/337/6101/1505 (2012).
8. Ludwig, J. *et al.* Neighborhoods, Obesity, and Diabetes – A Randomized Social Experiment. *New England Journal of Medicine* **365.** PMID: 22010917, 1509–1519. eprint: http://dx.doi.org/10.1056/NEJMsa1103216. http://dx.doi.org/10.1056/NEJMsa1103216 (2011).
9. Chetty, R., Hendren, N. & Katz, L. F. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review* **106,** 855–902. https://www.aeaweb.org/articles?id=10.1257/aer.20150572 (2016).
10. Galor, O. & Zeira, J. Income Distribution and Macroeconomics. *Review of Economic Studies* **60,** 35–52. http://ideas.repec.org/a/bla/restud/v60y1993i1p35-52.html (1993).
11. Bardhan, P., Bowles, S. & Gintis, H. in *Handbook of Income Distribution* (eds Atkinson, A. & Bourguignon, F.) (Elsevier-Science, North-Holland, 2000).
12. Carter, M. R. & Barrett, C. The Economics of Poverty Traps and Persistent Poverty: An Asset-based Approach. *Journal of Development Studies* (2006).
13. Galor, O., Moav, O. & Vollrath, D. Inequality in Land Ownership, the Emergence of Human Capital Promoting Institutions, and the Great Divergence. *Review of Economic Studies* **76,** 143 –179 (2009).
14. Currie, J. & Gahvari, F. Transfers in cash and in-kind: Theory meets the data. *Journal of Economic Literature* **46,** 333–383 (2008).

15. Aizer, A., Eli, S., Ferrie, J. & Lleras-Muney, A. The long-run impact of cash transfers to poor families. *American Economic Review* **106,** 935–971 (2016).

16. Parker, S. W. & Todd, P. E. Conditional Cash Transfers: The Case of Progresa/Oportunidades. *Journal of Economic Literature* **55,** 866–915. https://www.aeaweb.org/articles?id=10.1257/jel.20151233 (2017).

17. Barr, A., Eggleston, J. & Smith, A. A. Investing in Infants: the Lasting Effects of Cash Transfers to New Families. *The Quarterly Journal of Economics* **137,** 2539–2583. ISSN: 0033-5533. eprint: https://academic.oup.com/qje/article-pdf/137/4/2539/45946976/qjac023.pdf. https://doi.org/10.1093/qje/qjac023 (Apr. 2022).

18. Baird, S., De Hoop, J. & Özler, B. Income shocks and adolescent mental health. *Journal of Human Resources* **48,** 370–403 (2013).

19. Bastian, J. & Michelmore, K. The long-term impact of the earned income tax credit on children's education and employment outcomes. *Journal of Labor Economics* **36,** 1127–1163 (2018).

20. Chyn, E. Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review* **108,** 3028–3056 (2018).

21. De Mel, S., McKenzie, D. & Woodruff, C. One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka. *Science* **335,** 962–966 (2012).

22. Keswell, M. & Carter, M. R. Poverty and land redistribution. *Journal of Development Economics* **110,** 250–261. ISSN: 0304-3878. http://www.sciencedirect.com/science/article/pii/S0304387813001466 (2014).

23. Banerjee, A. *et al.* A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science* **348.** ISSN: 0036-8075. eprint: http://science.sciencemag.org/content/348/6236/1260799.full.pdf. http://science.sciencemag.org/content/348/6236/1260799 (2015).

24. Haushofer, J. & Shapiro, J. The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. *The Quarterly Journal of Economics* **131,** 1973. eprint: /oup/backfile/content_public/journal/qje/131/4/10.1093_qje_qjw025/5/qjw025.pdf. +http://dx.doi.org/10.1093/qje/qjw025 (2016).

25. Cunha, J. M., De Giorgi, G. & Jayachandran, S. The Price Effects of Cash Versus In-Kind Transfers. *The Review of Economic Studies* **86,** 240–281. ISSN: 0034-6527. eprint: https://academic.oup.com/restud/article-pdf/86/1/240/27285274/rdy018.pdf. https://doi.org/10.1093/restud/rdy018 (Apr. 2018).

26. Egger, D., Haushofer, J., Miguel, E., Niehaus, P. & Walker, M. General Equilibrium Effects of Cash Transfers: Experimental Evidence From Kenya. *Econometrica* **90,** 2603–2643. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA17945. https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17945 (2022).

27. Ozer, E. J., Fernald, L. C. H., Weber, A., Flynn, E. P. & VanderWeele, T. J. Does alleviating poverty affect mothers' depressive symptoms? A quasi-experimental investigation of Mexico's Oportunidades programme. *International Journal of Epidemiology* **40,** 1565–1576. ISSN: 03005771 (2011).

28. Kilburn, K., Thirumurthy, H., Halpern, C. T., Pettifor, A. & Handa, S. *Effects of a Large-Scale Unconditional Cash Transfer Program on Mental Health Outcomes of Young People in Kenya* 2015.

29. Walker, C. *Landmarked: land claims and land restitution in South Africa* (Ohio University Press, 2014).

30. Mani, A., Mullainathan, S., Shafir, E. & Zhao, J. Poverty Impedes Cognitive Function. *Science* **341,** 976–980. ISSN: 0036-8075. eprint: http://science.sciencemag.org/content/341/6149/976.full.pdf. http://science.sciencemag.org/content/341/6149/976 (2013).

31. Shah, A. K., Mullainathan, S. & Shafir, E. Some Consequences of Having Too Little. *Science* **338,** 682–685. ISSN: 0036-8075. eprint: http://science.sciencemag.org/content/338/6107/682.full.pdf. http://science.sciencemag.org/content/338/6107/682 (2012).

32. Blair, C. *et al.* Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child Development,* 1970–1984 (2011).

33. Haushofer, J. & Fehr, E. On the psychology of poverty. *Science* **344,** 862–867. ISSN: 0036-8075. eprint: http://science.sciencemag.org/content/344/6186/862.full.pdf. http://science.sciencemag.org/content/344/6186/862 (2014).

34. Koechlin, E. & Summerfield, C. An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences* **11,** 229–235. ISSN: 13646613 (2007).

35. Cooper, R. P. Complementary perspectives on cognitive control. *Topics in Cognitive Science* **3,** 208–211. ISSN: 17568757 (2011).

36. Badre, D. Defining an ontology of cognitive control requires attention to component interactions. *Topics in Cognitive Science* **3,** 217–221. ISSN: 17568757 (2011).

37. Fan, J. An information theory account of cognitive control. *Frontiers in Human Neuroscience* **8,** 1–16. ISSN: 1662-5161 (2014).

38. Hubbard, N. A. *et al.* Depressive thoughts limit working memory capacity in dysphoria. *Cognition and Emotion* **0,** 1–17. ISSN: 0269-9931 (2016).

39. Shanmugan, S. *et al.* Common and Dissociable Mechanisms of Executive System Dysfunction Across Psychiatric Disorders in Youth. *The American journal of psychiatry* **173,** 517–526. https://pubmed.ncbi.nlm.nih.gov/26806874 (May 2016).

40. Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement* **1,** 385–401 (1977).

41. Lund, C. *et al.* Poverty and mental disorders: breaking the cycle in low-income and middle-income countries. *The Lancet* **378,** 1502–1514. http://dx.doi.org/10.1016/S0140-6736(11)60754-X.

42. Engle, R. W. Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science* **11,** 19–23. eprint: http://cdp.sagepub.com/content/11/1/19.full.pdf+html. http://cdp.sagepub.com/content/11/1/19.abstract (2002).

43. Bilker, W. B. *et al.* Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment* **19,** 354–69. ISSN: 1552-3489. http://www.ncbi.nlm.nih.gov/pubmed/22605785 (2012).

44. Kaller, C. P., Unterrainer, J. M. & Stahl, C. Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment* **24,** 46–53. ISSN: 1040-3590 (2012).

45. Humes, G. E., Welsh, M. C., Retzlaff, P & Cookson, N. Towers of Hanoi and London: Reliability and validity of two executive function tasks. *Assessment* **4,** 249–257. ISSN: 1073-1911. <GotoISI>://A1997XU61700005 (1997).

46. Imbens, G. W. The Role of the Propensity Score in Estimating Dose-Response Functions. *Biometrica* **87,** 706–710 (2000).

47. Hirano, K. & Imbens, G. in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (eds Gelman, A. & Meng, X.-L.) 73–84 (Wiley, West Sussex, UK., 1973).

48. Imai, K. & van Dyk, D. A. Causal Inference With General Treatment Regimes: Generalizing the Propensity Score. *Journal of the American Statistical Association* **99,** 854–866. http://ideas.repec.org/a/bes/jnlasa/v99y2004p854-866.html (2004).

49. Hahn, J. On the role of the propensity score in efficient semiparamentric estimation of average treatment effects. *Econometrica* **66,** 315–331 (1998).

50. Bardhan, P. & Mookherjee, D. Determinants of Redistributive Politics: An Empirical Analysis of Land Reforms in West Bengal, India. *American Economic Review* **100,** 1572–1600. http://ideas.repec.org/a/aea/aecrev/v100y2010i4p1572-1600.html (2010).

51. Binswanger, H. P., Deininger, K. & Feder, G. in *Handbook of Development Economics* (Elsevier-Science, North-Holland: Amsterdam, 1995).

52. Banerjee, A., Gertler, P. & Ghatak, M. Empowerment and Efficiency: Tenancy Reform in West Bengal. *Journal of Political Economy* **110,** 239–280 (2002).

53. Besley, T. & Burgess, R. Land Reform, Poverty Reduction, and Growth: Evidence from India. *Quarterly Journal of Economics* **115,** 389–430 (2000).

54. Banerjee, A. & Iyer, L. History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *American Economic Review* **95,** 1190–1213. http://ideas.repec.org/a/aea/aecrev/v95y2005i4p1190-1213.html (2005).

55. Goldstein, M. & Udry, C. The Profits of Power: Land Rights and Agricultural Investment in Ghana. *Journal of Political Economy* **116,** 981–1022. http://ideas.repec.org/a/ucp/jpolec/v116y2008i6p981-1022.html (2008).

56. Eyal, K. & Burns, J. The parent trap: Cash transfers and the intergenerational transmission of depressive symptoms in South Africa. *World Development* **117,** 211–229. ISSN: 0305-750X. https://www.sciencedirect.com/science/article/pii/S0305750X19300208 (2019).

57. Loewenstein, G., Weber, E., Hsee, C. & Welch, N. Risk As Feelings. *Psychological bulletin* **127,** 267–86 (Mar. 2001).

58. Bia, M. & Mattei, A. A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score. *Stata Journal* **8,** 354–373. http://ideas.repec.org/a/tsj/stataj/v8y2008i3p354-373.html (2008).

59. Silverman, B. *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986).

60. Heckman, J. J. & Todd, P. E. A note on adapting propensity score matching and selection models to choice based samples. *Econometrics Journal* **12,** S230–S234. http://ideas.repec.org/a/ect/emjrnl/v12y2009is1ps230-s234.html (Jan. 2009).

61. Heckman, J. J. & Robb, R. J. Alternative methods for evaluating the impact of interventions : An overview. *Journal of Econometrics* **30,** 239–267. http://ideas.repec.org/a/eee/econom/v30y1985i1-2p239-267.html (1985).

62. Chamberlain, G. & Leamer, E. E. Matrix Weighted Averages and Posterior Bounds. *Journal of the Royal Statistical Society* **Series B,** 73–84 (1976).

63. Angrist, J. Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants. *Econometrica* **66,** 249–288 (1998).

64. Behrman, J., Cheng, Y. & Todd, P. Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach. *Review of Economics and Statistics* **86,** 108–32 (2004).

Appendix A

Further tables and figures

Table A.1: Minimum detectable effects for CES-D10 depression score (standard deviation increase)

Intra-cluster correlation	358 claims		500 claims	
	6	7	6	7
0	0.12	0.12	0.11	0.10
0.05	0.13	0.12	0.11	0.10
0.1	0.13	0.12	0.11	0.10
0.15	0.13	0.13	0.11	0.11
0.2	0.14	0.13	0.11	0.11

Test size and power: we assume standard values of $\alpha = 0.05$ and $\beta = 0.8$. Mean depression score for control group households estimated from Phase 1 of data collection is 9.04, with a standard deviation of 5.51.Two sample ranges are considered: a smaller sample of 6 or 7 CSM households each for 356 claims ($n = 2136-2492$); and a larger sample of 6 or 7 CSM households each for 500 claims ($n = 3000 - 3500$). ρ denotes the intra-cluster correlation (ICC). The ICC is a measure of the extent to which households within any given claim can be expected to have outcomes that are similar. For instance $\rho \approx 1$ means that every household in a claim has virtually the same level of consumption. By contrast $\rho \approx 0$ means that there is virtually no correlation between the consumption levels of households in a claim. We have allowed for a wide range of ICC values, with an upper limit of 0.20. Estimates represents minimum detectable effects (MDEs). The MDE is an estimate of the smallest level of impact that the applicable design is capable of detecting. For instance, the value of 0.11 reported in the last column of row 5 of the table indicates that an impact (reduction in depression scores) of 0.11 standard deviations, and no smaller, can be detected with a sample of $n = 7 \times 500 = 3500$ households, where the ICC is set to 0.20.

Table A.2: Minimum detectable effects for working memory (proportion correct increase)

Intra-cluster correlation	358 claims		500 claims	
	6	7	6	7
0	0.06	0.06	0.05	0.05
0.05	0.07	0.06	0.06	0.05
0.1	0.07	0.07	0.06	0.06
0.15	0.08	0.08	0.07	0.06
0.2	0.08	0.08	0.07	0.07

Test size and power: we assume standard values of $\alpha = 0.05$ and $\beta = 0.8$. Mean score as a proportion for control group households estimated from Phase 1 of data collection is 0.39. Two sample ranges are considered: a smaller sample of 6 or 7 CSM households each for 356 claims ($n = 2136 - 2492$); and a larger sample of 6 or 7 CSM households each for 500 claims ($n = 3000 - 3500$). ρ denotes the intra-cluster correlation (ICC). The ICC is a measure of the extent to which households within any given claim can be expected to have outcomes that are similar. For instance $\rho \approx 1$ means that every household in a claim has virtually the same level of consumption. By contrast $\rho \approx 0$ means that there is virtually no correlation between the consumption levels of households in a claim. We have allowed for a wide range of ICC values, with an upper limit of 0.20. Estimates represents minimum detectable effects (MDEs). The MDE is an estimate of the smallest level of impact that the applicable design is capable of detecting. For instance, the value of 0.07 reported in the last column of row 5 of the table indicates that an impact (increase in proportion correct) of 0.07, and no smaller, can be detected with a sample of $n = 7 \times 500 = 3500$ households, where the ICC is set to 0.20.

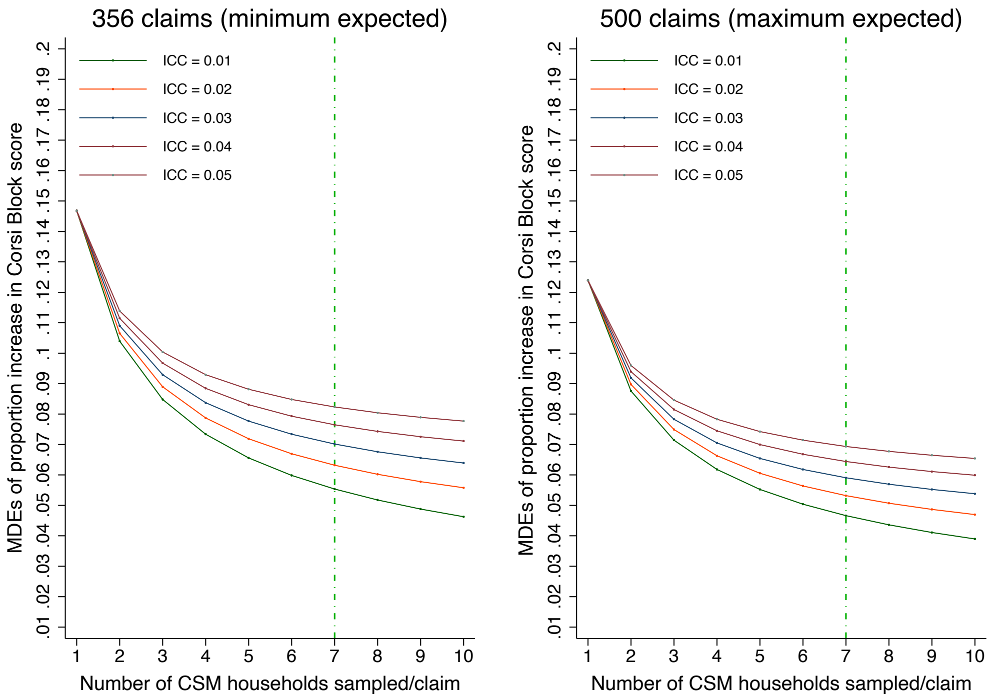


Figure A.1: Minimum detectable effects for working memory score (proportion correct increase)

Test size and power: we assume standard values of $\alpha = 0.05$ and $\beta = 0.8$. Mean score as a proportion for control group households estimated from Phase 1 of data collection is 0.39. Two sample ranges are considered: a smaller sample of $1 - 10$ CSM households each for 356 claims ($n = 356 - 3560$); and a larger sample of $1 - 10$ CSM households each for 500 claims ($n = 500 - 5000$). ICC denotes the intra-cluster correlation. Estimates represents minimum detectable effects (MDEs).

Table A.3: Summary statistics: individual level

	Control	Treatment	Total
CES-D-10 depression score (out of 30)	8.626	8.532	8.576
	(5.420)	(6.096)	(5.786)
Working memory span (standard deviations)	5.214	6.079	5.744
	(2.005)	(2.449)	(2.325)
Award made in 2013	.	0.0333	0.0333
	(.)	(0.180)	(0.180)
Award made in 2014	.	0.120	0.120
	(.)	(0.325)	(0.325)
Award made in 2015	.	0.0623	0.0623
	(.)	(0.242)	(0.242)
Award made in 2016	.	0.131	0.131
	(.)	(0.337)	(0.337)
Award made in 2017	.	0.176	0.176
	(.)	(0.381)	(0.381)
Award made in 2018	.	0.220	0.220
	(.)	(0.415)	(0.415)
Award made in 2019	.	0.0888	0.0888
	(.)	(0.285)	(0.285)
Award made in 2020	.	0.0777	0.0777
	(.)	(0.268)	(0.268)
Award made in 2021	.	0.0592	0.0592
	(.)	(0.236)	(0.236)
Award made in 2022	.	0.0321	0.0321
	(.)	(0.176)	(0.176)
Household size	5.072	4.879	4.970
	(3.114)	(2.876)	(2.992)
Opposition control in ward of household: 2011-2015	0	0.0890	0.0471
	(0)	(0.285)	(0.212)
Opposition control in ward of household: 2016-2020	0	0.283	0.150
	(0)	(0.451)	(0.357)
Opposition control in ward of dispossessed property: 2011-2015	0	0.147	0.0779
	(0)	(0.354)	(0.268)
Opposition control in ward of dispossessed property: 2016-2020	0	0.448	0.237
	(0)	(0.497)	(0.425)
Opposition control in ward of dispossessed property: 2021-	0.224	0.158	0.189
	(0.417)	(0.365)	(0.391)
Race (1= African)	0.768	0.674	0.719
	(0.422)	(0.469)	(0.450)
Race (1= Coloured)	0.213	0.258	0.237
	(0.409)	(0.438)	(0.425)
Race (1= Indian)	0.0149	0.0384	0.0273
	(0.121)	(0.192)	(0.163)
At least 1 English speaker in household	0.168	0.449	0.317
	(0.374)	(0.498)	(0.465)
At least 1 employed person in household	0.506	0.474	0.489
	(0.500)	(0.499)	(0.500)
Access to government	0.102	0.250	0.180
	(0.303)	(0.433)	(0.385)
Access to a political party	0.105	0.259	0.187
	(0.307)	(0.438)	(0.390)
Access to legal advice	0.155	0.489	0.332
	(0.362)	(0.500)	(0.471)
Gender is male	0.470	0.449	0.459
	(0.499)	(0.498)	(0.498)
Education (years)	10.19	10.75	10.49
	(3.547)	(3.708)	(3.644)
Real value of award (x1000)	.	39.31	39.31
	(.)	(110.9)	(110.9)
Observations	3735		

Access to government and political parties is defined as whether the household has at least one member who themselves are members of government or a political party, or whether they know someone who is. Access to legal advice is defined as whether the household has at least one person capable of providing legal advice or whether they know someone who can. The variable *Opposition control in ward of dispossessed property in 2011* is a dummy variable that measures whether the claim was finalised in the years between the 2011 and 2016 municipal elections and the dispossessed property was in a ward where opposition parties won the majority of seats in 2011. Similarly the 2016, and 2021 versions of this variable indicate claims finalised between 2016 and 2020, and 2021-2022 respectively and where opposition parties won the majority of seats in 2016 and 2021.

Table A.4: Maximum likelihood estimates of the determinants of award size (household level)

Variables	Model 1	Model 2
Race (1= African)	0.0313 (0.16)	-0.163 (-0.81)
Race (1= Coloured)	-0.520** (-2.68)	-0.584** (-3.05)
Race (1= Indian)	0.740** (3.04)	0.634** (2.64)
At least 1 English speaker in household	-0.371*** (-4.72)	-0.352*** (-4.57)
At least 1 employed person in household	-0.226** (-3.22)	-0.242*** (-3.52)
Access to government	0.167 (1.91)	0.149 (1.75)
Access to a political party	-0.120 (-1.38)	-0.135 (-1.59)
Access to legal advice	-0.00875 (-0.12)	0.00155 (0.02)
Award made in 2014		0.246 (1.14)
Award made in 2015		0.00372 (0.02)
Award made in 2016		-0.413 (-1.57)
Award made in 2017		-0.219 (-0.88)
Award made in 2018		-0.391 (-1.58)
Award made in 2019		0.511 (1.94)
Award made in 2020		0.458 (1.70)
Award made in 2021		-0.490 (-1.36)
Award made in 2022		-0.0435 (-0.14)
Household size		0.00174 (0.13)
Opposition control in ward of household: 2011		-0.305 (-1.92)
Opposition control in ward of household: 2016		-0.0351 (-0.35)
Opposition control in ward of dispossessed property: 2011		-0.338 (-1.67)
Opposition control in ward of dispossessed property: 2016		-0.225* (-2.17)
Opposition control in ward of dispossessed property: 2021		0.244 (0.85)
Constant	3.012*** (15.00)	3.470*** (11.09)
Observations	1574	1564

Access to government and political parties is defined as whether the household has at least one member who themselves are members of government or a political party, or whether they know someone who is. Access to legal advice is defined as whether the household has at least one person capable of providing legal advice or whether they know someone who can. The variable *Opposition control in ward of dispossessed property in 2011* is a dummy variable that measures whether the claim was finalised in the years between the 2011 and 2016 municipal elections and the dispossessed property was in a ward where opposition parties won the majority of seats in 2011. Similarly the 2016, and 2021 versions of this variable indicate claims finalised between 2016 and 2020, and 2021-2022 respectively and where opposition parties won the majority of seats in 2016 and 2021.

Table A.5: Maximum likelihood estimates of the determinants of award size (individual level)

Variables	Model 1	Model 2
eq1		
Award made in 2014	0.243	0.201
	(1.13)	(0.94)
Award made in 2015	0.00496	-0.0329
	(0.02)	(-0.14)
Award made in 2016	-0.414	-0.474
	(-1.58)	(-1.82)
Award made in 2017	-0.233	-0.295
	(-0.94)	(-1.19)
Award made in 2018	-0.403	-0.449
	(-1.63)	(-1.83)
Award made in 2019	0.512	0.459
	(1.95)	(1.75)
Award made in 2020	0.459	0.414
	(1.72)	(1.56)
Award made in 2021	-0.492	-0.531
	(-1.37)	(-1.49)
Award made in 2022	-0.0445	-0.0866
	(-0.15)	(-0.29)
Household size	0.00187	-0.00238
	(0.14)	(-0.18)
Opposition control in ward of household: 2011-2015	-0.303	-0.279
	(-1.92)	(-1.77)
Opposition control in ward of household: 2016-2020	-0.0357	-0.0419
	(-0.36)	(-0.42)
Opposition control in ward of dispossessed property: 2011-2015	-0.343	-0.396*
	(-1.71)	(-1.98)
Opposition control in ward of dispossessed property: 2016-2020	-0.226*	-0.222*
	(-2.18)	(-2.16)
Opposition control in ward of dispossessed property: 2021-	0.240	0.239
	(0.84)	(0.84)
Race (1= African)	-0.168	-0.189
	(-0.84)	(-0.95)
Race (1= Coloured)	-0.604**	-0.626***
	(-3.17)	(-3.30)
Race (1= Indian)	0.630**	0.592*
	(2.63)	(2.48)
At least 1 English speaker in household	-0.356***	-0.310***
	(-4.65)	(-3.91)
At least 1 employed person in household	-0.245***	-0.229**
	(-3.58)	(-3.29)
Access to government	0.136	0.148
	(1.61)	(1.76)
Access to a political party	-0.133	-0.132
	(-1.57)	(-1.58)
Access to legal advice	-0.000690	0.00681
	(-0.01)	(0.09)
Gender is male		0.178**
		(2.59)
Education (years)		-0.0324**
		(-3.04)
Constant	3.487***	3.822***
	(11.19)	(11.51)
eq2		
Constant	1.328***	1.321***
	(56.12)	(56.12)
Observations	1575	1575

Access to government and political parties is defined as whether the household has at least one member who themselves are members of government or a political party, or whether they know someone who is. Access to legal advice is defined as whether the household has at least one person capable of providing legal advice or whether they know someone who can. The variable *Opposition control in ward of dispossessed property in 2011* is a dummy variable that measures whether the claim was finalised in the years between the 2011 and 2016 municipal elections and the dispossessed property was in a ward where opposition parties won the majority of seats in 2011. Similarly the 2016, and 2021 versions of this variable indicate claims finalised between 2016 and 2020, and 2021-2022 respectively and where opposition parties won the majority of seats in 2016 and 2021.

Table A.6: Bayes factor tests of equality of conditional covariate means (household level)

	Normalized Treatment Intervals		
Variable	[0.09, 7.79]	[7.85, 25.50]	[25.54, 919.19]
Award made in 2014	10,702	10.47	9,3446
Award made in 2015	7,4159	10,129	10,641
Award made in 2016	9,9732	1,6702	5,9165
Award made in 2017	9,2303	5,3757	11.116
Award made in 2018	8,1361	10,228	4,6917
Award made in 2019	7,7931	5,5002	1,0284
Award made in 2020	9,0268	6,8691	11,009
Award made in 2021	10,201	2,4565	9,2629
Award made in 2022	10,416	10,524	11,107
Household size	7,3297	1,7179	8,8057
Opposition control in ward of household: 2011	11,245	9.978	9,6365
Opposition control in ward of household: 2016	9.961	6,4491	9,3702
Opposition control in ward of dispossessed property: 2011	11,178	7,4367	5,7145
Opposition control in ward of dispossessed property: 2016	7,8993	1,7708	10,758
Opposition control in ward of dispossessed property: 2021	9,3755	3,0414	8,8107
Race (1= African)	9,6652	6.408	6,4097
Race (1= Coloured	8,7315	1,6991	3,5083
Race (1= Indian)	3,2877	10,543	3,9535
At least 1 English speaker in household	7,4473	6,3982	4,2633
At least 1 employed person in household	9,9421	9.76	9,9183
Access to government	11,163	8,0951	9,1987
Access to a political party	11,262	10,485	9,6165
Access to legal advice	8,7953	10,069	9,9049

(a) The GPS is the predicted award size from a regression of award size on variables that could potential play a confounding role. To test that the GPS balances the data we partition the support of $\mathcal{D}$ into three mutually exclusive intervals, denoted as $G_1, \ldots, G_3$. Within each treatment interval G_k, we then compute the GPS $r(d_{G_k}, \mathbf{x}_i)$ at the mean of the interval $d_{G_k} \in G_k$. Then, for each of the three intervals we estimate the GPS at these treatment interval means d_{G_k} and then discretise the distribution of the GPS evaluated at this representative point.

(b) In our model, we chose 4 mutually exclusive blocks (5 in the case of the Model 2 from Table 4.3, denoted by $B_1^{(k)}, \ldots, B_4^{(k)}$. Within each interval $B_j^{(k)}$ for $j = 1, \ldots, 4$ and $k = 1, 2, 3$, we compute the difference in means for each covariate across different treatment intervals, but in the same GPS interval (i.e., j is held constant while k is varied). This results in 4 mean differences for each $d_{G_k} \in G_k$. This information is then collapsed into a single metric, by taking a weighted average of the differences at each representative point, where the weights are equivalent to the number of observations within each block $B_j^{(k)}$. This procedure is repeated for each covariate. In a final step, these weighted averages are then used to construct test statistics [47, 58].

Table A.7: Bayes factor tests of equality of conditional covariate means (individual level)

	Normalized Treatment Intervals		
Variable	[0.09, 7.79]	[7.85, 25.50]	[25.54, 919.19]
Award made in 2014	12,587	11,605	12,677
Award made in 2015	10,275	11,336	12,893
Award made in 2016	11,859	2,5755	7,3659
Award made in 2017	10,109	8,6251	12,431
Award made in 2018	11,223	12,216	6,1958
Award made in 2019	10,354	6,5085	1,9463
Award made in 2020	10,243	8,2485	11,028
Award made in 2021	11,874	2,7262	8.297
Award made in 2022	10,982	12,082	12,985
Household size	9,1062	2,1201	11,684
Opposition control in ward of household: 2011	12,542	10,995	12,592
Opposition control in ward of household: 2016	10,125	5,9127	10,449
Opposition control in ward of dispossessed property: 2011	13.18	9,2584	9.276
Opposition control in ward of dispossessed property: 2016	11,447	1,5522	11,992
Opposition control in ward of dispossessed property: 2021	10,373	3,1052	7,9756
Race (1= African)”	11,389	5,5314	10,595
Race (1= Coloured)”	10,252	1,0188	8,8238
Race (1= Indian)”	4,4354	12,214	7,4407
At least 1 English speaker in household	7,7357	8,7066	10,846
At least 1 employed person in household	11,082	11,884	12,661
Access to government	12,995	10,084	11,052
Access to a political party	13,212	12,171	10,278
Access to legal advice	10,286	11.25	11,174
Beneficiary is male	12,768	8,4987	12,139
Educaion (years)	7,7904	11,038	10,824

(a) The GPS is the predicted award size from a regression of award size on variables that could potential play a confounding role. To test that the GPS balances the data we partition the support of $\mathcal{D}$ into three mutually exclusive intervals, denoted as $G_1, \ldots, G_3$. Within each treatment interval G_k, we then compute the GPS $r(d_{G_k}, \mathbf{x}_i)$ at the mean of the interval $d_{G_k} \in G_k$. Then, for each of the three intervals we estimate the GPS at these treatment interval means d_{G_k} and then discretise the distribution of the GPS evaluated at this representative point.

(b) In our model, we chose 4 mutually exclusive blocks (5 in the case of the Model 2 from Table 4.3, denoted by $B_1^{(k)}, \ldots, B_4^{(k)}$. Within each interval $B_j^{(k)}$ for $j = 1, \ldots, 4$ and $k = 1, 2, 3$, we compute the difference in means for each covariate across different treatment intervals, but in the same GPS interval (i.e., j is held constant while k is varied). This results in 4 mean differences for each $d_{G_k} \in G_k$. This information is then collapsed into a single metric, by taking a weighted average of the differences at each representative point, where the weights are equivalent to the number of observations within each block $B_j^{(k)}$. This procedure is repeated for each covariate. In a final step, these weighted averages are then used to construct test statistics [47, 58].

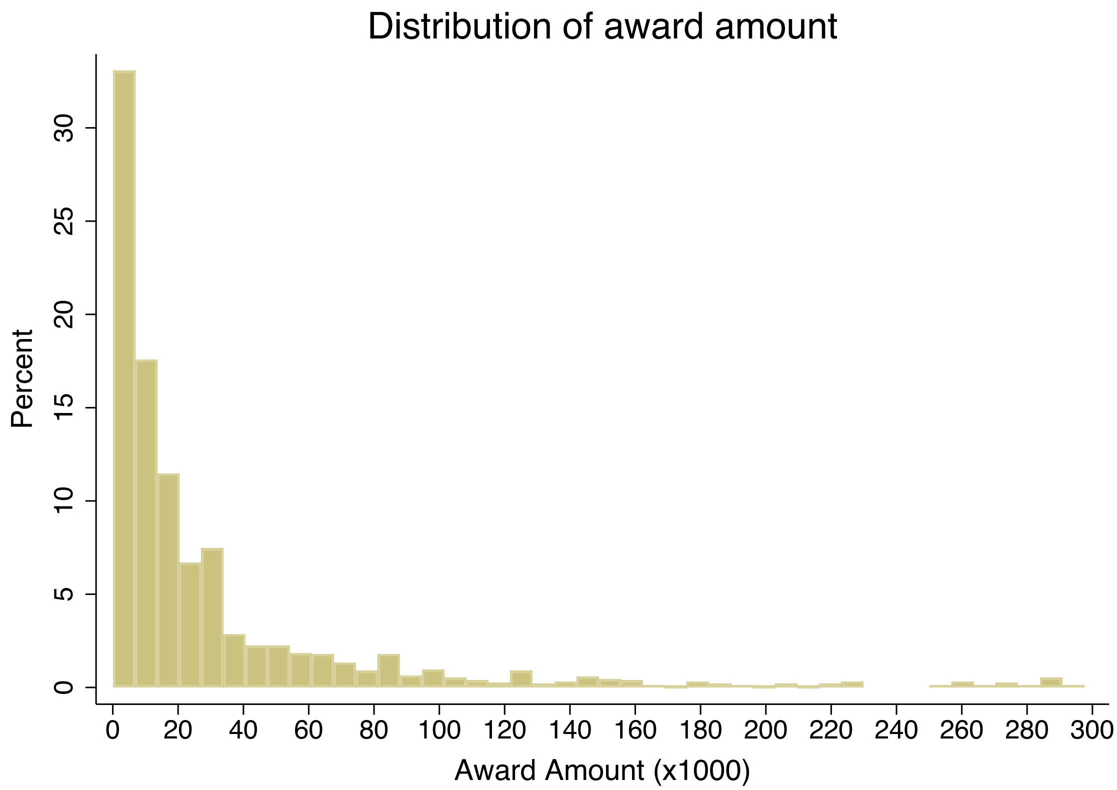


Figure A.2: Distribution of restitution awards

Appendix B

Statistical methodology: technical details

B.1 Propensity scores

To define an average treatment effect on the treated, we start by denoting S_p as the region of common support of p_i between the $D = 1$ and $D = 0$ distributions. Let N_1 denote the set of treated units, and let N_0 denote the set of control units. Now denote as n_1 the number of treated units falling into the common support region of the estimated propensity score density; i.e., the number of units falling into the set $N_1 \cap S_p$. Our matching estimator is then given by

$$\begin{aligned}
\delta &= (n_1)^{-1} \sum_{i \in N_1 \cap S_p} \Big(y_{1i} - \hat{\mathrm{E}}(y_{0i} | D_i = 1, p_i) \Big) \\
&= (n_1)^{-1} \sum_{i \in N_1 \cap S_p} \left(y_{1i} - \sum_{j \in N_0} \omega(i,j) y_{0j} \right) \qquad \text{(B.1)}
\end{aligned}$$

where $i \in N_1 \cap S_p$ denotes the ith treated unit from the set of units with common support on p_i. The second term in this expression serves as a matched substitute for the outcomes of a randomized-out unit of the treatment group, where the imputed counterfactual outcome $\sum_{j \in N_0} \omega(i,j) y_{0j}$ is a kernel-weighted average over the set of possible matches, with weight function:

$$\omega(i,j) = K\left(\frac{\mathbf{x}_j'\boldsymbol{\beta} - \mathbf{x}_i'\boldsymbol{\beta}}{h_n}\right) \Big/ \sum_{k \in N_0} K\left(\frac{\mathbf{x}_k'\boldsymbol{\beta} - \mathbf{x}_i'\boldsymbol{\beta}}{h_n}\right) \qquad \text{(B.2)}$$

where K is a kernel function, h_n is a bandwidth parameter and $\mathbf{x}_i'\boldsymbol{\beta}$ is the log of odds ratio.[1]

To estimate the propensity score, the dependent variable takes a value of 1 if a unit is in the treatment group, and a value of 0 if it is in the control group. A functional form for K and bandwidth choice must also be specified [59]. It is also common practice to not match directly on the predicted probability but rather on the log of the odds ratio. This is useful when the true sampling weights are unknown [60].

[1] It is usual to match directly on p_i. However, here we match on the log-odds ratio, for reasons we discuss below. Recall that $p_i = G(\mathbf{x}_i'\boldsymbol{\beta}) = e^{\mathbf{x}'\boldsymbol{\beta}}/(1 + e^{\mathbf{x}'\boldsymbol{\beta}})$, and $1 - p_i = 1/(1 + e^{\mathbf{x}_i'\boldsymbol{\beta}})$. Thus $\ln(p/(1-p)) = \mathbf{x}_i'\boldsymbol{\beta}$.

B.2 Generalised propensity scores

An average treatment effect on the treated is a data-weighted average of zero or negative impacts (likely to be experienced by recipients of very small awards) combined with positive impacts (likely to be experienced by recipients of larger awards). To explore this type of treatment heterogeneity, a natural starting point could be a random coefficients model [61]. However, if treatment status in non-linear in beneficiary characteristics, then the minimum mean square error approximation to the underlying conditional expectation function (CEF) does not have a straightforward interpretation, because the regression coefficients in such a model would actually represent a matrix-weighted average of the gradient of the CEF [62]. This problem can be overcome if the CEF of the continuous treatment variable (award size) is restricted to be linear [63]. However we would still require a large number of observations for each award size to justify this approach. An alternative approach that does not necessitate such an assumption is to used a generalised propensity score approach [47].

We begin by restricting attention to the sample of units in the treatment group, $i \in N_1$. We then postulate an award-response function $y_i(d)$ for all $d \in \mathcal{D}$ given that $i \in N_1$; i.e., each unit could have any potential outcome from the set $\mathcal{D}$ depending on the award size received. When treatment status is binary, we have $\mathcal{D} = \{0, 1\}$, but here we let $\mathcal{D} = \{d_0, d_1\}$.

The evaluation problem of course results from the fact that each unit realises exactly one outcome; that associated with the actual award received, $y_i = y_i(D_i)$, where $D_i \in [d_0, d_1]$. However, under the continuous treatment case, the problem is further complicated by the fact that there is more than one possible counterfactual award. We therefore define the impact of restitution in this continuous case in terms of an average award-response function, $\mu(d) = E[y_i(d)]$. Our goal then is to uncover non-constant treatment effects by taking the difference between this average and some benchmark level of treatment:

$$\theta(d) = \mu(d) - \mu(\tilde{d}) = E[y(d)] - E[y(\tilde{d})] \qquad \tilde{d}, d \in \mathcal{D} \tag{B.3}$$

where $\tilde{d}$ serves as the benchmark award.[2] In our empirical estimates, d is a chosen level of award from its distribution that we wish to calculate a treatment effect for and $\tilde{d}$ is d shifted by an arbitrary amount, which we set to 1. For instance we might choose 10 different cut points of the award distribution between 10000 and 100000. Then there are 10 value of d corresponding to $\{10000 \ldots 100000\}$ and correspondingly 10 different values of $\tilde{d} \in \{11000, 21000, \ldots, 91000, 100100\}$.

As in the binary approach, valid identification depends on an independence assumption regarding treatment assignment. Weak unconfoundedness is then defined as:[3]

$$y(d) \perp D | \mathbf{x} \qquad \forall d \in \mathcal{D}$$

To fix ideas, define $r(d, x)$ as the conditional density of treatment award given the covariates

$$r(d, x) = f_{D|\mathbf{x}}(d, x) \tag{B.4}$$

and define a generalized propensity score (GPS) $R = r(D, X)$. Using this framework, it can then be shown that assignment to treatment *intensity* (i.e., how large an award was made to a beneficiary) is unconfounded

[2] To simplify the notation, we drop the i subscripting when making reference to realised outcomes or treatment levels.

[3] This is essentially a weaker version of the"strong ignorability" assumption, generalised to multi-valued treatments [46, 47]. This assumption does not require joint independence of all potential outcomes, $\{y(d)\}_{t \in [t_0, t_1]}$ but rather that conditional independence holds for each value of D. Alternative conceptualisations are used sometimes [48, 64].

when $f_D(d|r(d,x),Y(d)) = f_D(d|r(d,X))$ [47]. Unconfoundedness is of course trivially met if intensity of treatment is a random process unrelated to expected impacts of the treatment. In section 4.2.2 we explain why this is probably not a unreasonable assumption in the case of the restitution programme.

Under the assumption of unconfoundedness, the GPS can be used to identify $\mu(d)$. Two steps are involved in showing why this is the case. First, in estimating the conditional expectation of the outcomes, all relevant information about the conditional density of the treatment is controlled for by directly conditioning on the treatment level D and the GPS $\widehat{R}_i$. Second, to estimate the award-response function, $\beta(d, r(d,X))$, at a particular level of the treatment, we average this conditional expectation over the GPS at that particular level of the treatment, $\mu(d) = \mathrm{E}[\beta(d, r(d,X))]$ and then by iterated expectations, $\mathrm{E}[\beta(d, r(d,\mathbf{x}))] = \mathrm{E}[\mathrm{E}[y(d) \mid r(d,\mathbf{x})]] = \mathrm{E}[y(d)]$ obtains.[4] Thus knowledge of $\beta(D,R)$ will identify the average award-response function, under weak unconfoundedness, conditional on the GPS.

To implement this estimator, we assume that the conditional density of the award of treatment is normally distributed with

$$D_i|\mathbf{x}_i \sim N(\mathbf{x}_i\boldsymbol{\beta}, \sigma^2) \tag{B.5}$$

These parameters are then estimated by maximum-likelihood, and the estimated GPS recovered as:

$$\widehat{R}_i = \frac{1}{\sqrt{2\pi\widehat{\sigma}^2}} \exp\left\{ -\frac{1}{2\widehat{\sigma}^2}\left(D_i - \widehat{\mathbf{x}_i\boldsymbol{\beta}}\right)^2 \right\} \tag{B.6}$$

To estimate the award-response function, we model the conditional expectation of y_i, as a flexible function of D_i and R_i

$$\begin{aligned} \beta(D_i, R_i) = E[Y_i|D_i, \widehat{R}_i] \quad &= \quad \alpha_0 + \alpha_1 D_i + \alpha_2 D_i^2 + \alpha_3 \widehat{R}_i + \\ & \quad\quad \alpha_4 \widehat{R}_i^2 + \alpha_5 D_i \widehat{R}_i \end{aligned} \tag{B.7}$$

Equation B.7 is then estimated by OLS.[5] Once we have the estimated α parameters, we can then recover the average award response function $\mathrm{E}[y(d)]$. Recall that $\mathrm{E}[y(d)]$ is identified for particular levels of award, so the average must be taken over all units at award level d. This effectively equates to averaging over the GPS for each award level d. By changing the treatment level at which the averaging takes place, we recover an estimate of the entire award-response function.[6] This gives a treatment effect estimator of the form

$$\begin{aligned} \widehat{\psi}(d) = \widehat{E[Y(d)]} \quad &= \quad \frac{1}{n_1}\sum_{i=1}^{n_1}(\widehat{\alpha}_0 + \widehat{\alpha}_1 \cdot d + \widehat{\alpha}_2 \cdot d^2 + \widehat{\alpha}_3 \cdot \widehat{r}(d,\mathbf{x}_i) \\ & \quad\quad + \widehat{\alpha}_4 \cdot \widehat{r}(d,\mathbf{x}_i)^2 + \widehat{\alpha}_5 \cdot d \cdot \widehat{r}(d,\mathbf{x}_i)) \end{aligned} \tag{B.8}$$

[4] Importantly, note that under this approach, the averaging that is used to construct $\mu(d)$ takes places over the GPS score evaluated at the treatment level of interest, $r(d,\mathbf{x})$, and not over the GPS itself.

[5] It should be stressed that the regression function $\beta(d,r)$ does not have a causal interpretation. In particular, the derivative with respect to the treatment level d does not represent an average effect of changing the level of treatment for any particular subpopulation. We also experimented with various specifications for this regression and conclude that not much additional explanatory power is added by including higher than second-order polynomials in D and $\widehat{R}$.

[6] We estimate standard errors and confidence intervals for each point along the award-response function using bootstrap methods. However, in principal, analytical standard errors can also be computed given the parametric forms of the GPS and $\beta(D,R)$.

Appendix C

Land restitution approval pipeline

C.1 Lodgement

The first step is the lodgement of a claim. When a claim has passed through this step, it will have been been registered, screened and validated as eligible and legal (meaning it has been established that the dispossession did indeed take place as evidenced by the historical record, and was governed by one or more race-based Acts of Parliament passed between 1913-1994).

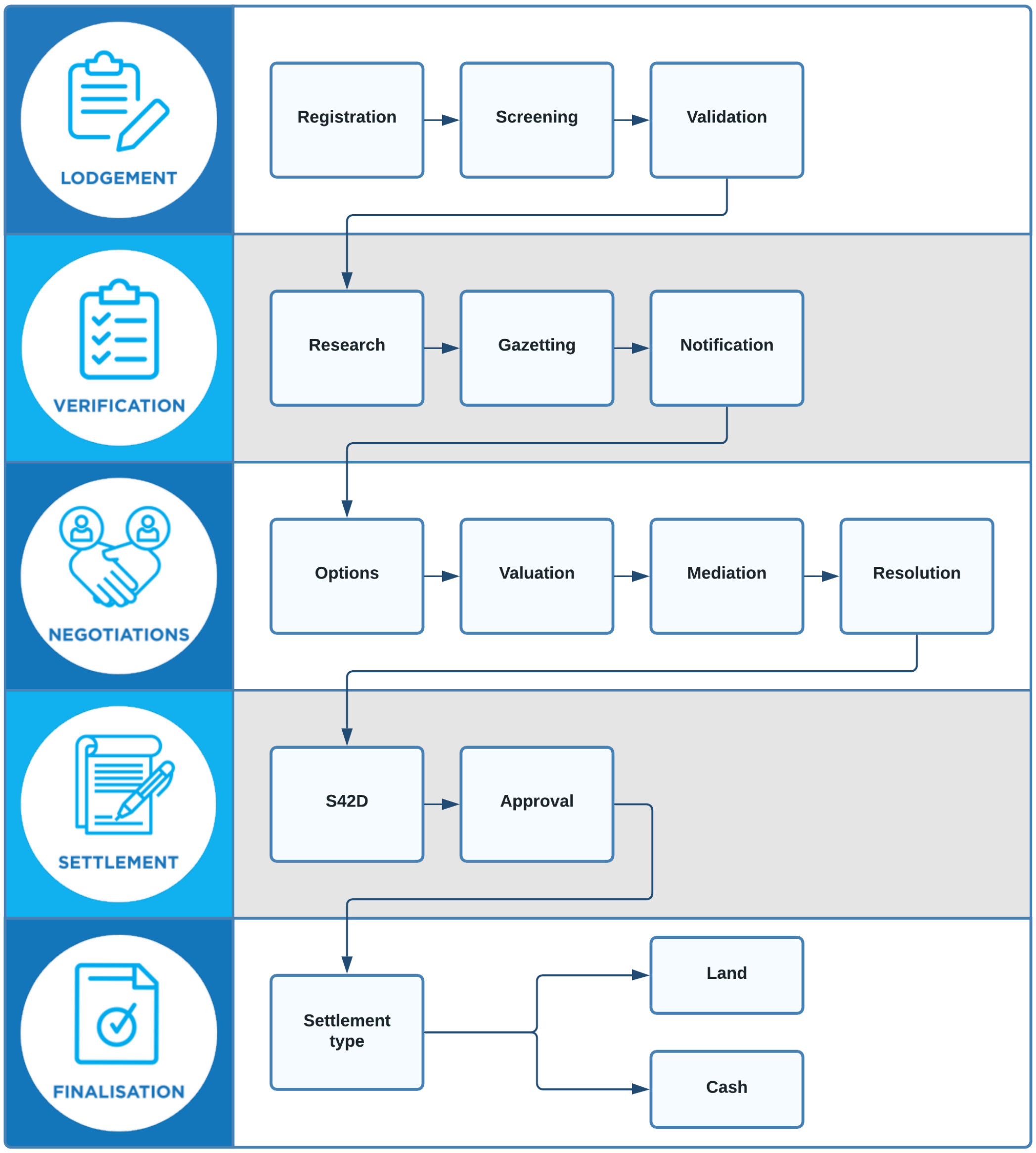
C.2 Verification

The next step in the process is Verification. Detailed forensic work is carried out during this step, both on eligibility in terms of the Restitution Act, as well as verification that a right to the originally confiscated land did indeed vest with an individual that can be linked (dynastically or otherwise) to the individual lodging the claim. Teams of attorneys and advocates employed by the Commission for the Restitution of Land Rights (CRLR) (hereafter, the "Commission"') do the bulk of this forensic work, which culminates in what is known as the "Rule 5 Research Report," where the "rule 5" refers to that part of the Act that mandates what evidentiary standard must met for a claim to be verified and how that evidence should be documented. This process of "research" leads to official public notice of the claim being given through publication in the Government Gazette.

C.3 Negotiations

Once a claim has been validated as eligible and legal, a process of negotiation ensues between the Commission and the claimants. Most often the Commission plays a mediating role between multiple beneficiaries, since there will almost always be several beneficiaries linked to one claim. This is a key step in the process and is where claimants and beneficiaries come together to discuss how they'd like to be compensated. This step in the process, called the "Options" workshop, is how the Restitution Act gives expression to the principles of local participation and active citizenry. It is the key mechanism provided by the Act for community engagement and collective decision making. The process is iterative. The beneficiaries must decide first on a "settlement option;" i.e., whether they prefer financial compensation or a land award. The next step in

Figure C.1: Land restitution claims settlement approval process



the process puts a value to the settlement option and this is taken back to the beneficiaries who must then reach a negotiated position.[1]

C.4 Settlement and finalisation

Once the negotiations process reaches a conclusion, a legal agreement known as the Section 42 D agreement is prepared in accordance with Section 42 D of the Restitution Act (hereafter "S42D"). Once the S42D agreement has been signed, the claim is considered approved ("Settled") in the legal sense. Funds are then committed to the claim administratively through a ledger system with a target window set for the disbursement of monies or land rights to the beneficiaries. This penultimate step presents an administrative queue. At some point during this target window, the committed funds are spent, either through financial transfers to beneficiaries, or the registration of the land title deed in the names of the beneficiaries. When this final step is concluded, the claim is considered "Finalised."

[1] Disputes that arise during the negotiations phase are handled by the Land Claims Court (LLC). The LLC is a special court set up as part of the restitution process to adjudicate only on the issue of land contestations. A dispute that is taken before the LLC can be resolved either through a judgment or court order (in the instance that the parties eventually come to a resolution). At this stage the claim is considered a legally "settled" matter. However in rare occasions when a judgment is issued, it can be further challenged in a High Court.

Appendix D

Consumption Aggregation

Table D.1: Expenditure items within goods categories

Goods category	Items included
Food	Mealie meal, samp, flour and bread, rice, pasta, breakfast cereal and porridge, baby food and baby formula, biscuits, cakes, rusks, sugar, jam, honey, chocolates and sweets, soft drinks and juices, red meat, canned red meat, chicken, fresh fish and shell fish and tinned fish, dried peas, lentils and beans, fruit and vegetables, potatoes, other vegetables, fruits and nuts, tinned fruit and vegetables and soya products, oil for cooking, margarine, butter, ghee, other fats, peanut butter, milk, cheese, yoghurt, dried milk, eggs, salt, spices, coffee, tea, food hampers, readymade meals, meals prepared outside the home (incl. restaurants and take-aways) and other food expenditure
Alcohol and tobacco	Cigarettes and tobacco, beer, wine and spirits
Hygiene	Personal care and cleaning agents
Transport	Transport
Fuel	Petrol, oil and car
Healthcare	Medical professionals, hospital fees, medical supplies, traditional healers, homeopaths etc and health insurance
Schooling	Educational policies, school fees and tuition, school books and, uniforms and other school expenses
Services and Maintenance	Water, electricity, other energy sources, municipal rates, levies, home maintenance and swimming pool maintenance
Other non-food goods	Entertainment, sport, jewellery and watches, literary materials, cell phone expenses, telephone expenses, lotto, gambling and horse-racing, internet, trips and holidays, ceremonies, car payments, life insurance, funeral policies, short-term insurance, kitchen equipment, linen, fabric, hire purchase payments, furniture, appliances, childcare, membership dues, domestic labour, pets, toys, gifts and lobola
Clothing and shoes	Shoes, clothes, account payments and clothing fabric

The table shows the expenditure items that are included in each goods category. All expenditure amounts are in December 2022 rands. For food items, expenditure is the reported value of food *purchases* in the last month. Where a non-food item is reported as monthly expenditure (items within alcohol and tobacco, transport, fuel, and entertainment and hygiene categories) the goods category is calculated as the sum across values of items.